Dell Configuration Guide for the S6000 System

9.5(0.1)

Notes, Cautions, and Warnings

NOTE: A NOTE indicates important information that helps you make better use of your computer.

CAUTION: A CAUTION indicates either potential damage to hardware or loss of data and tells you how to avoid the problem.

WARNING: A WARNING indicates a potential for property damage, personal injury, or death.

Copyright © 2014 Dell Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. Dell™ and the Dell logo are trademarks of Dell Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

2014 - 07

Rev. A00

Contents

1 About this Guide.. 32

- Audience.. 32
- Conventions.. 32
- Related Documents.. 32

2 Configuration Fundamentals.. 33

- Accessing the Command Line.. 33
- CLI Modes.. 33
  - Navigating CLI Modes.. 34
- The do Command.. 37
- Undoing Commands.. 38
- Obtaining Help.. 39
- Entering and Editing Commands.. 39
- Command History.. 40
- Filtering show Command Outputs.. 40
- Multiple Users in Configuration Mode.. 42

3 Getting Started.. 43

- Console Access.. 43
  - Serial Console.. 43
- Default Configuration.. 44
- Configuring a Host Name.. 44
- Accessing the System Remotely.. 45
  - Accessing the and S6000Remotely.. 45
  - Configure the Management Port IP Address.. 45
  - Configure a Management Route.. 46
  - Configuring a Username and Password.. 46
- Configuring the Enable Password.. 46
- Configuration File Management.. 47
  - Copy Files to and from the System.. 47
  - Save the Running-Configuration.. 48
  - Configure the Overload Bit for a Startup Scenario.. 49
  - Viewing Files.. 49
- Managing the File System.. 50
- View Command History.. 51
- Upgrading Dell Networking OS.. 51
- Using HTTP for File Transfers.. 51
- Using Hashes to Validate Software Images.. 52

4 Management... 54
- Configuring Privilege Levels... 54
  - Creating a Custom Privilege Level...54
  - Removing a Command from EXEC Mode.. 54
  - Moving a Command from EXEC Privilege Mode to EXEC Mode..................................... 54
  - Allowing Access to CONFIGURATION Mode Commands.. 55
  - Allowing Access to the Following Modes...55
  - Applying a Privilege Level to a Username...57
  - Applying a Privilege Level to a Terminal Line... 57
- Configuring Logging... 57
  - Audit and Security Logs...58
  - Configuring Logging Format 59
  - Setting Up a Secure Connection to a Syslog Server...60
- Log Messages in the Internal Buffer... 61
  - Configuration Task List for System Log Management... 61
- Disabling System Logging... 61
- Sending System Messages to a Syslog Server...62
  - Configuring a UNIX System as a Syslog Server.. 62
- Changing System Logging Settings..62
- Display the Logging Buffer and the Logging Configuration.. 63
- Configuring a UNIX Logging Facility Level...64
- Synchronizing Log Messages... 65
- Enabling Timestamp on Syslog Messages... 65
- File Transfer Services... 66
  - Configuration Task List for File Transfer Services... 66
  - Enabling the FTP Server.. 66
  - Configuring FTP Server Parameters...67
  - Configuring FTP Client Parameters.. 67
- Terminal Lines.. 68
  - Denying and Permitting Access to a Terminal Line...68
  - Configuring Login Authentication for Terminal Lines... 69
- Setting Time Out of EXEC Privilege Mode... 70
- Using Telnet to get to Another Network Device.. 70
- Lock CONFIGURATION Mode.. 71
  - Viewing the Configuration Lock Status..71
- Restoring the Factory Default Settings...72
  - S6000MXL Switch...72
  - Important Points to Remember... 72
  - Restoring Factory Default Environment Variables.. 73

5 802.1X.. 75

The Port-Authentication Process....................76
  EAP over RADIUS....................78
Configuring 802.1X....................78
  Related Configuration Tasks....................78
Important Points to Remember....................79
Enabling 802.1X....................79
Configuring Request Identity Re-Transmissions....................81
  Configuring a Quiet Period after a Failed Authentication....................81
Forcibly Authorizing or Unauthorizing a Port....................82
Re-Authenticating a Port....................83
Configuring Timeouts....................84
Configuring Dynamic VLAN Assignment with Port Authentication....................85
Guest and Authentication-Fail VLANs....................86
  Configuring a Guest VLAN....................87
  Configuring an Authentication-Fail VLAN....................87

6 Access Control Lists (ACLs)....................89

IP Access Control Lists (ACLs)....................90
  CAM Usage....................90
  Implementing ACLs on Dell Networking OS....................92
Important Points to Remember....................93
  Configuration Task List for Route Maps....................93
  Configuring Match Routes....................96
  Configuring Set Conditions....................97
  Configure a Route Map for Route Redistribution....................98
  Configure a Route Map for Route Tagging....................99
  Continue Clause....................99
IP Fragment Handling....................100
  IP Fragments ACL Examples....................100
  Layer 4 ACL Rules Examples....................101
Configure a Standard IP ACL....................102
  Configuring a Standard IP ACL Filter....................103
Configure an Extended IP ACL....................104
  Configuring Filters with a Sequence Number....................104
  Configuring Filters Without a Sequence Number....................105
Configure Layer 2 and Layer 3 ACLs....................106
Assign an IP ACL to an Interface....................107
Applying an IP ACL....................107
  Counting ACL Hits....................108
Configure Ingress ACLs....................108
Configure Egress ACLs....................109
  Applying Egress Layer 3 ACLs (Control-Plane)....................110

IP Prefix Lists....110
  Implementation Information....111
  Configuration Task List for Prefix Lists....111
ACL Resequencing....115
  Resequencing an ACL or Prefix List....115
Route Maps....117
  Implementation Information....117
Flow-Based Monitoring Support for ACLs....117
  Behavior of Flow-Based Monitoring....118
  Enabling Flow-Based Monitoring....119

7 Bidirectional Forwarding Detection (BFD)....121
How BFD Works....121
  BFD Packet Format....122
  BFD Sessions....124
  BFD Three-Way Handshake....124
  Session State Changes....125
Important Points to Remember....126
Configure BFD....126
  Configure BFD for Physical Ports....127
  Configure BFD for Static Routes....130
  Configure BFD for OSPF....132
  Configure BFD for OSPFv3....135
  Configure BFD for IS-IS....136
  Configure BFD for BGP....139
  Configure BFD for VRRP....146
  Configuring Protocol Liveness....149
  Troubleshooting BFD....149

8 Border Gateway Protocol IPv4 (BGPv4)....151
Autonomous Systems (AS)....151
Sessions and Peers....153
  Establish a Session....154
Route Reflectors....154
BGP Attributes....155
  Best Path Selection Criteria....156
  Weight....158
  Local Preference....158
  Multi-Exit Discriminators (MEDs)....159
  Origin....160
  AS Path....161
  Next Hop....161

Multiprotocol BGP..162
Implement BGP with Dell Networking OS...162
  Additional Path (Add-Path) Support.. 162
  Advertise IGP Cost as MED for Redistributed Routes.. 162
  Ignore Router-ID for Some Best-Path Calculations...163
  Four-Byte AS Numbers.. 163
  AS4 Number Representation.. 164
  AS Number Migration...165
  BGP4 Management Information Base (MIB)..167
  Important Points to Remember..167
Configuration Information...168
BGP Configuration..168
  Enabling BGP...169
  Configuring AS4 Number Representations.. 173
  Configuring Peer Groups.. 175
  Configuring BGP Fast Fall-Over...177
  Configuring Passive Peering... 179
  Maintaining Existing AS Numbers During an AS Migration...180
  Allowing an AS Number to Appear in its Own AS Path... 181
  Enabling Graceful Restart...182
  Enabling Neighbor Graceful Restart..183
  Filtering on an AS-Path Attribute...183
  Regular Expressions as Filters.. 185
  Redistributing Routes..186
  Enabling Additional Paths...187
  Configuring IP Community Lists...187
  Configuring an IP Extended Community List...189
  Filtering Routes with Community Lists... 190
  Manipulating the COMMUNITY Attribute..190
  Changing MED Attributes..192
  Changing the LOCAL_PREFERENCE Attribute...192
  Changing the NEXT_HOP Attribute..193
  Changing the WEIGHT Attribute...194
  Enabling Multipath.. 194
  Filtering BGP Routes... 194
  Filtering BGP Routes Using Route Maps... 196
  Filtering BGP Routes Using AS-PATH Information.. 197
  Configuring BGP Route Reflectors... 197
  Aggregating Routes...198
  Configuring BGP Confederations..199
  Enabling Route Flap Dampening..199
  Changing BGP Timers.. 202

    Enabling BGP Neighbor Soft-Reconfiguration....202
    Route Map Continue....204
  Enabling MBGP Configurations....204
  BGP Regular Expression Optimization....205
  Debugging BGP....205
    Storing Last and Bad PDUs....206
    Capturing PDUs....207
    PDU Counters....208
  Sample Configurations....209

9 Content Addressable Memory (CAM)....215
  CAM Allocation....215
  Test CAM Usage....217
  View CAM Profiles....217
  View CAM-ACL Settings....218
  View CAM Usage....220
  CAM Optimization....220
  Troubleshoot CAM Profiling....221
    CAM Profile Mismatches....221
    QoS CAM Region Limitation....221
    Syslog Error When the Table is Full....221
    Syslog Warning Upon 90 Percent Utilization of CAM....222
    Syslog Warning for Discrepancies Between Configured Extended Prefixes....222

10 Control Plane Policing (CoPP)....223
  Configure Control Plane Policing....224
    Configuring CoPP for Protocols....225
    Configuring CoPP for CPU Queues....227
    Show Commands....228

11 Data Center Bridging (DCB)....230
  Ethernet Enhancements in Data Center Bridging....230
    Priority-Based Flow Control....231
    Enhanced Transmission Selection....233
    Data Center Bridging Exchange Protocol (DCBx)....235
    Data Center Bridging in a Traffic Flow....235
  Enabling Data Center Bridging....236
  QoS dot1p Traffic Classification and Queue Assignment....236
  SNMP Support for PFC and Buffer Statistics Tracking....237
  Configuring Priority-Based Flow Control....238
    Configuring Lossless Queues....240
    Configuring the PFC Buffer in a Switch Stack....241

Configure Enhanced Transmission Selection....242
- ETS Prerequisites and Restrictions.... 242
- Creating a QoS DCB Output Policy.... 243
- Creating an ETS Priority Group.... 245
- Applying an ETS Output Policy for a Priority Group to an Interface.... 246
- ETS Operation with DCBx....247
- Configuring Bandwidth Allocation for DCBx CIN.... 248

Applying DCB Policies in a Switch Stack.... 248

Applying DCB Policies with an ETS Configuration....249

Configure a DCBx Operation.... 249
- DCBx Operation....250
- DCBx Port Roles....250
- DCB Configuration Exchange.... 252
- Configuration Source Election.... 252
- Propagation of DCB Information.... 253
- Auto-Detection and Manual Configuration of the DCBx Version.... 253
- Behavior of Tagged Packets.... 254
- Configuration Example for DSCP and PFC Priorities....254
- DCBx Example....255
- DCBx Prerequisites and Restrictions....256
- Configuring DCBx....256

Verifying the DCB Configuration.... 260

Generation of PFC for a Priority for Untagged Packets....271

Operations on Untagged Packets.... 272

Performing PFC Using DSCP Bits Instead of 802.1p Bits....272

PFC and ETS Configuration Examples.... 273

Using PFC and ETS to Manage Data Center Traffic....273
- PFC and ETS Configuration Command Examples....275
- Using PFC and ETS to Manage Converged Ethernet Traffic in a Switch Stack.... 276
- Hierarchical Scheduling in ETS Output Policies.... 276

Configuring DCB Maps and its Attributes....277
- DCB Map: Configuration Procedure.... 277
- Important Points to Remember....278
- Applying a DCB Map on a Port.... 278
- Configuring PFC without a DCB Map.... 279
- Configuring Lossless Queues....280

Data Center Bridging: Default Configuration.... 281

Configuring PFC and ETS in a DCB Map....281
- PFC Configuration Notes.... 281
- PFC Prerequisites and Restrictions....282
- ETS Configuration Notes....283
- ETS Prerequisites and Restrictions.... 283

Priority-Based Flow Control Using Dynamic Buffer Method....284
  Pause and Resume of Traffic....284
  Buffer Sizes for Lossless or PFC Packets....285
Interworking of DCB Map With DCB Buffer Threshold Settings....285
Configuring the Dynamic Buffer Method....286
Applying a DCB Map in a Switch Stack....288

12 Dynamic Host Configuration Protocol (DHCP)....289
DHCP Packet Format and Options....289
Assign an IP Address using DHCP....291
Implementation Information....292
Configure the System to be a DHCP Server....293
  Configuring the Server for Automatic Address Allocation....293
  Specifying a Default Gateway....295
  Configure a Method of Hostname Resolution....295
  Using DNS for Address Resolution....295
  Using NetBIOS WINS for Address Resolution....295
  Creating Manual Binding Entries....296
  Debugging the DHCP Server....296
  Using DHCP Clear Commands....296
Configure the System to be a DHCP Client....297
  Configuring the DHCP Client System....297
  DHCP Client on a Management Interface....299
  DHCP Client Operation with Other Features....299
Configure the System for User Port Stacking (Option 230)....300
Configure Secure DHCP....301
  Option 82....301
  DHCP Snooping....302
  Drop DHCP Packets on Snooped VLANs Only....304
  Dynamic ARP Inspection....304
  Configuring Dynamic ARP Inspection....305
Source Address Validation....306
  Enabling IP Source Address Validation....306
  DHCP MAC Source Address Validation....307
  Enabling IP+MAC Source Address Validation....307

13 Equal Cost Multi-Path (ECMP)....309
ECMP for Flow-Based Affinity....309
  Configuring the Hash Algorithm....309
  Enabling Deterministic ECMP Next Hop....309
  Configuring the Hash Algorithm Seed....310
Link Bundle Monitoring....310

Managing ECMP Group Paths.. 311
Creating an ECMP Group Bundle.. 312
Modifying the ECMP Group Threshold...312
Support for /128 IPv6 and /32 IPv4 Prefixes in Layer 3 Host Table and LPM Table................. 313
Support for ECMP in host table... 313
Support for moving /128 IPv6 Prefixes and /32 IPv4 Prefixes 314

14 FCoE Transit..315

Fibre Channel over Ethernet...315
Ensure Robustness in a Converged Ethernet Network.. 315
FIP Snooping on Ethernet Bridges..317
FIP Snooping in a Switch Stack..319
Using FIP Snooping.. 319
FIP Snooping Prerequisites.. 319
Important Points to Remember.. 320
Enabling the FCoE Transit Feature...320
Enable FIP Snooping on VLANs.. 321
Configure the FC-MAP Value..321
Configure a Port for a Bridge-to-Bridge Link... 321
Configure a Port for a Bridge-to-FCF Link..321
Impact on Other Software Features.. 321
FIP Snooping Restrictions..322
Configuring FIP Snooping..322
Displaying FIP Snooping Information..323
FCoE Transit Configuration Example...329

15 Flex Hash and Optimized Boot-Up...331

Flex Hash Capability Overview..331
Configuring the Flex Hash Mechanism... 331
Configuring Fast Boot and LACP Fast Switchover..332
Optimizing the Boot Time.. 332
Booting Process When Optimized Boot Time Mechanism is Enabled....................................333
Guidelines for Configuring Optimized Booting Mechanism...333
Interoperation of Applications with Fast Boot and System States..334
LACP and IPv4 Routing...334
LACP and IPv6 Routing... 335
BGP Graceful Restart.. 335
Cold Boot Caused by Power Cycling the System.. 335
Unexpected Reload of the System... 335
Software Upgrade.. 336
LACP Fast Switchover.. 336
Changes to BGP Multipath.. 336

Delayed Installation of ECMP Routes Into BGP....336
RDMA Over Converged Ethernet (RoCE) Overview....337
Preserving 802.1Q VLAN Tag Value for Lite Subinterfaces....338

16 Force10 Resilient Ring Protocol (FRRP)....339
Protocol Overview....339
Ring Status....340
Multiple FRRP Rings....341
Important FRRP Points....341
Important FRRP Concepts....341
Implementing FRRP....343
FRRP Configuration....343
Creating the FRRP Group....343
Configuring the Control VLAN....344
Configuring and Adding the Member VLANs....345
Setting the FRRP Timers....346
Clearing the FRRP Counters....347
Viewing the FRRP Configuration....347
Viewing the FRRP Information....347
Troubleshooting FRRP....348
Configuration Checks....348
Sample Configuration and Topology....348

17 GARP VLAN Registration Protocol (GVRP)....350
Important Points to Remember....350
Configure GVRP....351
Related Configuration Tasks....351
Enabling GVRP Globally....352
Enabling GVRP on a Layer 2 Interface....352
Configure GVRP Registration....352
Configure a GARP Timer....353

18 Internet Group Management Protocol (IGMP)....355
IGMP Implementation Information....355
IGMP Protocol Overview....355
IGMP Version 2....355
IGMP Version 3....357
Configure IGMP....360
Related Configuration Tasks....360
Viewing IGMP Enabled Interfaces....361
Selecting an IGMP Version....361
Viewing IGMP Groups....362

Adjusting Timers.......... 362
  Adjusting Query and Response Timers.......... 362
  Adjusting the IGMP Querier Timeout Value.......... 363
Configuring a Static IGMP Group.......... 363
Enabling IGMP Immediate-Leave.......... 364
IGMP Snooping.......... 364
  IGMP Snooping Implementation Information.......... 364
  Configuring IGMP Snooping.......... 364
  Removing a Group-Port Association.......... 365
  Disabling Multicast Flooding.......... 365
  Specifying a Port as Connected to a Multicast Router.......... 366
  Configuring the Switch as Querier.......... 366
Fast Convergence after MSTP Topology Changes.......... 367
Egress Interface Selection (EIS) for HTTP and IGMP Applications.......... 367
  Protocol Separation.......... 368
  Enabling and Disabling Management Egress Interface Selection.......... 369
  Handling of Management Route Configuration.......... 370
  Handling of Switch-Initiated Traffic.......... 370
  Handling of Switch-Destined Traffic.......... 371
  Handling of Transit Traffic (Traffic Separation).......... 372
  Mapping of Management Applications and Traffic Type.......... 372
  Behavior of Various Applications for Switch-Initiated Traffic 373
  Behavior of Various Applications for Switch-Destined Traffic 374
  Interworking of EIS With Various Applications.......... 375
Designating a Multicast Router Interface.......... 376

19 Interfaces.......... 377
Basic Interface Configuration.......... 377
Advanced Interface Configuration.......... 377
Interface Types.......... 378
View Basic Interface Information.......... 378
Enabling a Physical Interface.......... 380
Physical Interfaces.......... 380
  Configuration Task List for Physical Interfaces.......... 381
  40G to 1G Breakout Cable Adaptor.......... 381
  Overview of Layer Modes.......... 381
  Configuring Layer 2 (Data Link) Mode.......... 382
  Configuring Layer 2 (Interface) Mode.......... 382
  Configuring Layer 3 (Network) Mode.......... 383
  Configuring Layer 3 (Interface) Mode.......... 383
Egress Interface Selection (EIS).......... 384
  Important Points to Remember.......... 384

- Configuring EIS........385
Management Interfaces........385
- Configuring Management Interfaces........385
- Configuring Management Interfaces on the S-Series........386
VLAN Interfaces........386
Loopback Interfaces........387
Null Interfaces........388
Port Channel Interfaces........388
- Port Channel Definition and Standards........388
- Port Channel Benefits........389
- Port Channel Implementation........389
- 10/100/1000 Mbps Interfaces in Port Channels........389
- Configuration Tasks for Port Channel Interfaces........390
- Creating a Port Channel........390
- Adding a Physical Interface to a Port Channel........391
- Reassigning an Interface to a New Port Channel........393
- Configuring the Minimum Oper Up Links in a Port Channel........393
- 394
- Assigning an IP Address to a Port Channel........394
- Deleting or Disabling a Port Channel........395
- Load Balancing Through Port Channels........395
- Load-Balancing on the S- Series........395
- Changing the Hash Algorithm........396
Bulk Configuration........397
- Interface Range........397
- Bulk Configuration Examples........398
Defining Interface Range Macros........399
- Define the Interface Range........399
- Choosing an Interface-Range Macro........400
Monitoring and Maintaining Interfaces........400
- Maintenance Using TDR........401
Splitting QSFP Ports to SFP+ Ports........402
Converting a QSFP or QSFP+ Port to an SFP or SFP+ Port........402
- Important Points to Remember........403
- Support for LM4 Optics........404
- Example Scenarios........404
Link Dampening........408
- Important Points to Remember........408
- Enabling Link Dampening........408
Link Bundle Monitoring........410
Using Ethernet Pause Frames for Flow Control........410
- Threshold Settings........411

Enabling Pause Frames....411
Configure the MTU Size on an Interface....412
Port-Pipes....413
Auto-Negotiation on Ethernet Interfaces....413
Setting the Speed and Duplex Mode of Ethernet Interfaces....414
Set Auto-Negotiation Options....415
View Advanced Interface Information....416
Configuring the Interface Sampling Size....417
Dynamic Counters....418
Clearing Interface Counters....418
Enhanced Validation of Interface Ranges....419
Compressing Configuration Files....419

20 IPv4 Routing....423
IP Addresses....423
Implementation Information....423
Configuration Tasks for IP Addresses....423
Assigning IP Addresses to an Interface....424
Configuring Static Routes....425
Configure Static Routes for the Management Interface....426
Using the Configured Source IP Address in ICMP Messages....427
Configuring the ICMP Source Interface....427
Configuring the Duration to Establish a TCP Connection....427
Enabling Directed Broadcast....428
Resolution of Host Names....428
Enabling Dynamic Resolution of Host Names....429
Specifying the Local System Domain and a List of Domains....429
Configuring DNS with Traceroute....430
ARP....431
Configuration Tasks for ARP....431
Configuring Static ARP Entries....431
Enabling Proxy ARP....432
Clearing ARP Cache....432
ARP Learning via Gratuitous ARP....433
Enabling ARP Learning via Gratuitous ARP....433
ARP Learning via ARP Request....433
Configuring ARP Retries....434
ICMP....435
Configuration Tasks for ICMP....435
Enabling ICMP Unreachable Messages....435
UDP Helper....436
Configure UDP Helper....436

Important Points to Remember....436
Enabling UDP Helper....436
Configuring a Broadcast Address....437
Configurations Using UDP Helper....437
UDP Helper with Broadcast-All Addresses....437
UDP Helper with Subnet Broadcast Addresses....438
UDP Helper with Configured Broadcast Addresses....439
UDP Helper with No Configured Broadcast Addresses....439
Troubleshooting UDP Helper....440

21 IPv6 Routing....441
Protocol Overview....441
Extended Address Space....441
Stateless Autoconfiguration....441
IPv6 Headers....442
Longest Prefix Match (LPM) Table and IPv6 /65 – /128 support....442
IPv6 Header Fields....444
Extension Header Fields....445
Addressing....446
Implementing IPv6 with Dell Networking OS....448
ICMPv6....450
Path MTU Discovery....450
IPv6 Neighbor Discovery....451
IPv6 Neighbor Discovery of MTU Packets....452
Configuring the IPv6 Recursive DNS Server....452
Debugging IPv6 RDNSS Information Sent to the Host....453
Displaying IPv6 RDNSS Information....453
Secure Shell (SSH) Over an IPv6 Transport....454
Configuration Tasks for IPv6....454
Adjusting Your CAM-Profile....455
Assigning an IPv6 Address to an Interface....455
Assigning a Static IPv6 Route....456
Configuring Telnet with IPv6....457
SNMP over IPv6....457
Showing IPv6 Information....457
Showing an IPv6 Interface....458
Showing IPv6 Routes....459
Showing the Running-Configuration for an Interface....460
Clearing IPv6 Routes....460

22 iSCSI Optimization....462
iSCSI Optimization Overview....462

  - Monitoring iSCSI Traffic Flows....................464
  - Application of Quality of Service to iSCSI Traffic Flows....................464
  - Information Monitored in iSCSI Traffic Flows....................464
  - Detection and Auto-Configuration for Dell EqualLogic Arrays....................465
  - Configuring Detection and Ports for Dell Compellent Arrays....................466
  - Synchronizing iSCSI Sessions Learned on VLT-Lags with VLT-Peer....................466
  - Enable and Disable iSCSI Optimization....................466
- Default iSCSI Optimization Values....................467
- iSCSI Optimization Prerequisites....................468
- Configuring iSCSI Optimization....................468
- Displaying iSCSI Optimization Information....................470

23 Intermediate System to Intermediate System....................472

- IS-IS Protocol Overview....................472
- IS-IS Addressing....................472
- Multi-Topology IS-IS....................473
  - Transition Mode....................474
  - Interface Support....................474
  - Adjacencies....................474
- Graceful Restart....................474
  - Timers....................475
- Implementation Information....................475
- Configuration Information....................476
  - Configuration Tasks for IS-IS....................476
  - Configuring the Distance of a Route....................485
  - Changing the IS-Type....................485
  - Redistributing IPv4 Routes....................488
  - Redistributing IPv6 Routes....................489
  - Configuring Authentication Passwords....................490
  - Setting the Overload Bit....................490
  - Debugging IS-IS....................491
- IS-IS Metric Styles....................492
- Configure Metric Values....................492
  - Maximum Values in the Routing Table....................493
  - Change the IS-IS Metric Style in One Level Only....................493
  - Leaks from One Level to Another....................495
- Sample Configurations....................495

24 Link Aggregation Control Protocol (LACP)....................498

- Introduction to Dynamic LAGs and LACP....................498
  - Important Points to Remember....................498
  - LACP Modes....................499

    - Configuring LACP Commands....499
  - LACP Configuration Tasks....500
    - Creating a LAG....500
    - Configuring the LAG Interfaces as Dynamic....501
    - Setting the LACP Long Timeout....501
    - Monitoring and Debugging LACP....502
  - Shared LAG State Tracking....502
    - Configuring Shared LAG State Tracking....503
    - Important Points about Shared LAG State Tracking....504
  - LACP Basic Configuration Example....505
    - Configure a LAG on ALPHA....505

25 Layer 2....513

  - Manage the MAC Address Table....513
    - Clearing the MAC Address Table....513
    - Setting the Aging Time for Dynamic Entries....513
    - Configuring a Static MAC Address....514
    - Displaying the MAC Address Table....514
  - MAC Learning Limit....514
    - Setting the MAC Learning Limit....515
    - mac learning-limit Dynamic....515
    - mac learning-limit mac-address-sticky....516
    - mac learning-limit station-move....516
    - mac learning-limit no-station-move....516
    - Learning Limit Violation Actions....517
    - Setting Station Move Violation Actions....517
    - Recovering from Learning Limit and Station Move Violations....518
  - NIC Teaming....518
  - Configure Redundant Pairs....519
    - Important Points about Configuring Redundant Pairs....521
  - Far-End Failure Detection....522
    - FEFD State Changes....523
    - Configuring FEFD....524
    - Enabling FEFD on an Interface....525
    - Debugging FEFD....526

26 Link Layer Discovery Protocol (LLDP)....528

  - 802.1AB (LLDP) Overview....528
    - Protocol Data Units....528
  - Optional TLVs....529
    - Management TLVs....529
  - TIA-1057 (LLDP-MED) Overview....531

  - TIA Organizationally Specific TLVs....532
- Configure LLDP....536
  - Related Configuration Tasks....536
  - Important Points to Remember....537
  - LLDP Compatibility....537
- CONFIGURATION versus INTERFACE Configurations....537
- Enabling LLDP....538
  - Disabling and Undoing LLDP....538
- Enabling LLDP on Management Ports....538
  - Disabling and Undoing LLDP on Management Ports....538
- Advertising TLVs....539
- Viewing the LLDP Configuration....540
- Viewing Information Advertised by Adjacent LLDP Agents....541
- Configuring LLDPDU Intervals....542
- Configuring Transmit and Receive Mode....542
- Configuring a Time to Live....543
- Debugging LLDP....544
- Relevant Management Objects....545

27 Microsoft Network Load Balancing....551

- NLB Unicast Mode Scenario....551
- NLB Multicast Mode Scenario....552
- Limitations With Enabling NLB on Switches....552
- Benefits and Working of Microsoft Clustering....552
- Enable and Disable VLAN Flooding 553
- Configuring a Switch for NLB 553
  - 553

28 Multicast Source Discovery Protocol (MSDP)....554

- Protocol Overview....554
- Anycast RP....556
- Implementation Information....556
- Configure Multicast Source Discovery Protocol....556
  - Related Configuration Tasks....556
- Enable MSDP....560
- Manage the Source-Active Cache....561
  - Viewing the Source-Active Cache....561
  - Limiting the Source-Active Cache....562
  - Clearing the Source-Active Cache....562
  - Enabling the Rejected Source-Active Cache....562
- Accept Source-Active Messages that Fail the RFP Check....562
- Specifying Source-Active Messages....566

- Limiting the Source-Active Messages from a Peer....567
- Preventing MSDP from Caching a Local Source.... 567
- Preventing MSDP from Caching a Remote Source....568
- Preventing MSDP from Advertising a Local Source.... 569
- Logging Changes in Peership States....570
- Terminating a Peership....570
- Clearing Peer Statistics.... 570
- Debugging MSDP.... 571
- MSDP with Anycast RP.... 571
- Configuring Anycast RP.... 573
  - Reducing Source-Active Message Flooding.... 573
  - Specifying the RP Address Used in SA Messages....573
- MSDP Sample Configurations.... 576

29 Multiple Spanning Tree Protocol (MSTP).... 579

- Protocol Overview.... 579
- Spanning Tree Variations....580
  - Implementation Information....580
- Configure Multiple Spanning Tree Protocol....580
  - Related Configuration Tasks.... 580
- Enable Multiple Spanning Tree Globally.... 581
- Adding and Removing Interfaces.... 581
- Creating Multiple Spanning Tree Instances.... 581
- Influencing MSTP Root Selection....583
- Interoperate with Non-Dell Networking OS Bridges.... 583
- Changing the Region Name or Revision.... 584
- Modifying Global Parameters....584
- Modifying the Interface Parameters....585
- Configuring an EdgePort....586
- Flush MAC Addresses after a Topology Change.... 587
- MSTP Sample Configurations....587
  - Router 1 Running-ConfigurationRouter 2 Running-ConfigurationRouter 3 Running-ConfigurationSFTOS Example Running-Configuration.... 588
- Debugging and Verifying MSTP Configurations.... 591

30 Multicast Features....594

- Enabling IP Multicast.... 594
- Multicast with ECMP....594
- Implementation Information....595
- First Packet Forwarding for Lossless Multicast....596
- Multicast Policies.... 596
  - IPv4 Multicast Policies.... 596

31 Open Shortest Path First (OSPFv2 and OSPFv3)......................................605
Protocol Overview...605
Autonomous System (AS) Areas...605
Area Types...606
Networks and Neighbors...607
Router Types..607
Designated and Backup Designated Routers...609
Link-State Advertisements (LSAs)...609
Router Priority and Cost..611
OSPF with Dell Networking OS...611
Graceful Restart...612
Fast Convergence (OSPFv2, IPv4 Only)...613
Multi-Process OSPFv2 (IPv4 only)...613
RFC-2328 Compliant OSPF Flooding..614
OSPF ACK Packing...615
Setting OSPF Adjacency with Cisco Routers..615
Configuration Information...616
Configuration Task List for OSPFv2 (OSPF for IPv4)..616
Configuration Task List for OSPFv3 (OSPF for IPv6)..632
Enabling IPv6 Unicast Routing..633
Assigning IPv6 Addresses on an Interface..633
Assigning Area ID on an Interface...633
Assigning OSPFv3 Process ID and Router ID Globally...634
Configuring Stub Areas...634
Configuring Passive-Interface...634
Redistributing Routes...635
Configuring a Default Route..635
Enabling OSPFv3 Graceful Restart..636
OSPFv3 Authentication Using IPsec...638
Troubleshooting OSPFv3...645

32 Policy-based Routing (PBR)... 647
Overview... 647
Implementing Policy-based Routing with Dell Networking OS.......................................649
Configuration Task List for Policy-based Routing...649
PBR Exceptions (Permit)..652
Sample Configuration..654
Create the Redirect-List GOLDAssign Redirect-List GOLD to Interface 2/11View Redirect-List GOLD...655

33 PIM Sparse-Mode (PIM-SM)... 657

Implementation Information..657
Protocol Overview..657
Requesting Multicast Traffic..657
Refuse Multicast Traffic..658
Send Multicast Traffic..658
Configuring PIM-SM..659
Related Configuration Tasks..659
Enable PIM-SM..659
Configuring S,G Expiry Timers..660
Configuring a Static Rendezvous Point..661
Overriding Bootstrap Router Updates..662
Configuring a Designated Router..662
Creating Multicast Boundaries and Domains..663
Enabling PIM-SM Graceful Restart..663

34 PIM Source-Specific Mode (PIM-SSM)..664
Implementation Information..664
Important Points to Remember..664
Configure PIM-SMM..665
Related Configuration Tasks..665
Enabling PIM-SSM..665
Use PIM-SSM with IGMP Version 2 Hosts..665
Configuring PIM-SSM with IGMPv2..666

35 Port Monitoring..668
Important Points to Remember..668
Port Monitoring..669
Configuring Port Monitoring..670
Enabling Flow-Based Monitoring..672
Remote Port Mirroring..673
Remote Port Mirroring Example..673
Configuring Remote Port Mirroring..674
Displaying Remote-Port Mirroring Configurations..676
Configuring the Sample Remote Port Mirroring..676
Configuring the Encapsulated Remote Port Mirroring..680
Configuration steps for ERPM 680
ERPM Behavior on a typical Dell Networking OS 682
Decapsulation of ERPM packets at the Destination IP/ Analyzer..682

36 Per-VLAN Spanning Tree Plus (PVST+)..684
Protocol Overview..684
Implementation Information..685

Configure Per-VLAN Spanning Tree Plus....685
  Related Configuration Tasks....685
Enabling PVST+....685
Disabling PVST+....686
Influencing PVST+ Root Selection....686
Modifying Global PVST+ Parameters....688
Modifying Interface PVST+ Parameters....689
Configuring an EdgePort....690
PVST+ in Multi-Vendor Networks....691
Enabling PVST+ Extend System ID....691
PVST+ Sample Configurations....692

37 Quality of Service (QoS)....694
Implementation Information....696
Port-Based QoS Configurations....696
  Setting dot1p Priorities for Incoming Traffic....697
  Honoring dot1p Priorities on Ingress Traffic....697
  Configuring Port-Based Rate Policing....698
  Configuring Port-Based Rate Shaping....698
Policy-Based QoS Configurations....699
  Classify Traffic....700
  Dot1p to Queue Mapping Requirement....703
  Create a QoS Policy....704
  DSCP Color Maps....706
  Create Policy Maps....708
Enabling QoS Rate Adjustment....712
Enabling Strict-Priority Queueing....713
Queue Classification Requirements for PFC Functionality....713
Support for marking dot1p value in L3 Input Qos Policy....714
Weighted Random Early Detection....714
  Creating WRED Profiles....715
  Applying a WRED Profile to Traffic....716
  Displaying Default and Configured WRED Profiles....716
  Displaying WRED Drop Statistics....716
Pre-Calculating Available QoS CAM Space....716
Specifying Policy-Based Rate Shaping in Packets Per Second....717
Configuring Policy-Based Rate Shaping....718
Configuring Weights and ECN for WRED....719
  Global Service Pools With WRED and ECN Settings....720
Configuring WRED and ECN Attributes....721
Guidelines for Configuring ECN for Classifying and Color-Marking Packets....722
  Sample configuration to mark non-ecn packets as “yellow” with Multiple traffic class....723

  - Classifying Incoming Packets Using ECN and Color-Marking.....723
  - Sample configuration to mark non-ecn packets as “yellow” with single traffic class.....726
- Applying Layer 2 Match Criteria on a Layer 3 Interface.....727
  - Managing Hardware Buffer Statistics.....727
- Enabling Buffer Statistics Tracking 728

38 Routing Information Protocol (RIP)..... 729

- Protocol Overview.....729
  - RIPv1.....729
  - RIPv2.....729
- Implementation Information.....730
- Configuration Information.....730
  - Configuration Task List.....730
  - RIP Configuration Example.....737

39 Remote Monitoring (RMON).....743

- Implementation Information.....743
- Fault Recovery.....743
  - Setting the rmon Alarm.....744
  - Configuring an RMON Event.....745
  - Configuring RMON Collection Statistics.....745
  - Configuring the RMON Collection History.....746

40 Rapid Spanning Tree Protocol (RSTP).....747

- Protocol Overview.....747
- Configuring Rapid Spanning Tree.....747
  - Related Configuration Tasks.....747
- Important Points to Remember.....748
  - RSTP and VLT.....748
- Configuring Interfaces for Layer 2 Mode.....748
- Enabling Rapid Spanning Tree Protocol Globally.....749
- Adding and Removing Interfaces.....751
- Modifying Global Parameters.....752
  - Enabling SNMP Traps for Root Elections and Topology Changes.....753
- Modifying Interface Parameters.....753
- Enabling SNMP Traps for Root Elections and Topology Changes.....754
- Influencing RSTP Root Selection.....754
- Configuring an EdgePort.....754
- Configuring Fast Hellos for Link State Detection.....755

41 Software-Defined Networking (SDN)..... 757

42 Service Provider Bridging....758
VLAN Stacking....758
Important Points to Remember....759
Configure VLAN Stacking....760
Creating Access and Trunk Ports....760
Enable VLAN-Stacking for a VLAN....761
Configuring the Protocol Type Value for the Outer VLAN Tag....761
Configuring Dell Networking OS Options for Trunk Ports....761
Debugging VLAN Stacking....762
VLAN Stacking in Multi-Vendor Networks....763
VLAN Stacking Packet Drop Precedence....766
Enabling Drop Eligibility....766
Honoring the Incoming DEI Value....767
Marking Egress Packets with a DEI Value....768
Dynamic Mode CoS for VLAN Stacking....768
Mapping C-Tag to S-Tag dot1p Values....770
Layer 2 Protocol Tunneling....770
Implementation Information....772
Enabling Layer 2 Protocol Tunneling....773
Specifying a Destination MAC Address for BPDUs....773
Setting Rate-Limit BPDUs....773
Debugging Layer 2 Protocol Tunneling....774
Provider Backbone Bridging....774

43 sFlow....775
Overview....775
Implementation Information....775
Important Points to Remember....776
Enabling Extended sFlow....776
Enabling and Disabling sFlow on an Interface....777
sFlow Show Commands....777
Displaying Show sFlow Global....777
Displaying Show sFlow on an Interface....778
Displaying Show sFlow on a Stack-unit....778
Configuring Specify Collectors....779
Changing the Polling Intervals....779
Back-Off Mechanism....779
sFlow on LAG ports....780
Enabling Extended sFlow....780
Important Points to Remember....781

44 Simple Network Management Protocol (SNMP)....................................783
Protocol Overview...783
Implementation Information...783
SNMPv3 Compliance With FIPS...783
Configuration Task List for SNMP...785
Related Configuration Tasks..785
Important Points to Remember...785
Set up SNMP...785
Creating a Community...786
Setting Up User-Based Security (SNMPv3)..786
Reading Managed Object Values..787
Writing Managed Object Values..788
Configuring Contact and Location Information using SNMP.........................789
Subscribing to Managed Object Value Updates using SNMP.........................790
Enabling a Subset of SNMP Traps...791
Copy Configuration Files Using SNMP...793
Copying a Configuration File..795
Copying Configuration Files via SNMP..795
Copying the Startup-Config Files to the Running-Config.........................796
Copying the Startup-Config Files to the Server via FTP.............................796
Copying the Startup-Config Files to the Server via TFTP..........................797
Copy a Binary File to the Startup-Configuration..797
Additional MIB Objects to View Copy Statistics...798
Obtaining a Value for MIB Objects..798
Manage VLANs using SNMP...799
Creating a VLAN..799
Assigning a VLAN Alias...799
Displaying the Ports in a VLAN...800
Add Tagged and Untagged Ports to a VLAN..800
Managing Overload on Startup...801
Enabling and Disabling a Port using SNMP...801
Fetch Dynamic MAC Entries using SNMP..802
Deriving Interface Indices...804
Monitor Port-Channels..805
Troubleshooting SNMP Operation..806

45 Storm Control...807
Configure Storm Control...807
Configuring Storm Control from INTERFACE Mode...................................807
Configuring Storm Control from CONFIGURATION Mode.........................807

46 Spanning Tree Protocol (STP).. 808
Protocol Overview..808
Configure Spanning Tree...808
    Related Configuration Tasks...808
Important Points to Remember...809
Configuring Interfaces for Layer 2 Mode..809
Enabling Spanning Tree Protocol Globally..810
Adding an Interface to the Spanning Tree Group..812
Modifying Global Parameters..813
Modifying Interface STP Parameters..814
Enabling PortFast..814
    Prevent Network Disruptions with BPDU Guard...815
Selecting STP Root...817
STP Root Guard...818
    Root Guard Scenario...818
    Configuring Root Guard..819
Enabling SNMP Traps for Root Elections and Topology Changes.........................820
Configuring Spanning Trees as Hitless..820
STP Loop Guard..821
    Configuring Loop Guard...822
Displaying STP Guard Configuration..823

47 System Time and Date... 824
Network Time Protocol...824
    Protocol Overview...825
    Configure the Network Time Protocol..826
    Enabling NTP...826
    Setting the Hardware Clock with the Time Derived from NTP..............................826
    Configuring NTP Broadcasts..827
    Disabling NTP on an Interface...827
    Configuring a Source IP Address for NTP Packets..827
    Configuring NTP Authentication...828
Dell Networking OS Time and Date..831
    Configuration Task List 831
    Setting the Time and Date for the Switch Hardware Clock...................................831
    Setting the Time and Date for the Switch Software Clock.....................................831
    Setting the Timezone..832
    Set Daylight Saving Time..832
    Setting Daylight Saving Time Once...832
    Setting Recurring Daylight Saving Time..833

48 Tunneling 835
Configuring a Tunnel...835
Configuring Tunnel Keepalive Settings..836
Configuring a Tunnel Interface... 837
Configuring Tunnel allow-remote Decapsulation... 837
Configuring Tunnel source anylocal Decapsulation...838
Guidelines for Configuring Multipoint Receive-Only Tunnels...................................... 838
Multipoint Receive-Only Type and IP Unnumbered Interfaces for Tunnels.................839

49 Upgrade Procedures.. 840
Get Help with Upgrades..840

50 Virtual LANs (VLANs)...841
Default VLAN... 841
Port-Based VLANs.. 842
VLANs and Port Tagging...842
Configuration Task List...843
  Creating a Port-Based VLAN.. 843
  Assigning Interfaces to a VLAN.. 844
  Moving Untagged Interfaces... 845
  Assigning an IP Address to a VLAN...847
Configuring Native VLANs.. 847
Enabling Null VLAN as the Default VLAN...848

51 VLT Proxy Gateway...849
Proxy Gateway in VLT Domains..849
  LLDP organizational TLV for proxy gateway.. 851
  Sample Configuration Scenario for VLT Proxy Gateway...852
Configuring a Static VLT Proxy Gateway.. 854
Configuring an LLDP VLT Proxy Gateway.. 854

52 Virtual Link Trunking (VLT)...856
Overview.. 856
  VLT on Core Switches...857
  Enhanced VLT.. 857
VLT Terminology... 858
Configure Virtual Link Trunking.. 859
  Important Points to Remember... 859
  Configuration Notes...860
  Primary and Secondary VLT Peers... 864
  RSTP and VLT..864

VLT Bandwidth Monitoring....864
VLT and IGMP Snooping....865
VLT IPv6....865
VLT Port Delayed Restoration....865
PIM-Sparse Mode Support on VLT....866
VLT Routing 867
Non-VLT ARP Sync....870
RSTP Configuration....870
Preventing Forwarding Loops in a VLT Domain....870
Sample RSTP Configuration....871
Configuring VLT....871
eVLT Configuration Example....882
eVLT Configuration Step Examples....883
PIM-Sparse Mode Configuration Example....885
Verifying a VLT Configuration....886
Additional VLT Sample Configurations....889
Configuring Virtual Link Trunking (VLT Peer 1)Configuring Virtual Link Trunking (VLT Peer 2)Verifying a Port-Channel Connection to a VLT Domain (From an Attached Access Switch)....890
Troubleshooting VLT....892
Reconfiguring Stacked Switches as VLT....893
Specifying VLT Nodes in a PVLAN....894
Association of VLTi as a Member of a PVLAN....895
MAC Synchronization for VLT Nodes in a PVLAN....895
PVLAN Operations When One VLT Peer is Down....895
PVLAN Operations When a VLT Peer is Restarted....896
Interoperation of VLT Nodes in a PVLAN with ARP Requests....896
Scenarios for VLAN Membership and MAC Synchronization With VLT Nodes in PVLAN....896
Configuring a VLT VLAN or LAG in a PVLAN....898
Creating a VLT LAG or a VLT VLAN....898
Associating the VLT LAG or VLT VLAN in a PVLAN....899
Proxy ARP Capability on VLT Peer Nodes....900
Working of Proxy ARP for VLT Peer Nodes....901
VLT Nodes as Rendezvous Points for Multicast Resiliency....902

53 Virtual Extensible LAN (VXLAN)....903
Overview....903
Components of VXLAN network....903
Components of VXLAN network....903
Functional Overview of VXLAN Gateway....905
VXLAN Frame Format....905
Components of VXLAN Frame Format....906

Configuring and Controlling VXLAN from the NVP Controller GUI.....906
Configuring VxLAN Gateway..... 909
- Connecting to an NVP Controller.....909
- Advertising VXLAN Access Ports to Controller..... 910

Displaying VXLAN Configurations.....911

54 Virtual Routing and Forwarding (VRF)..... 913

VRF Overview..... 913
VRF Configuration Notes.....914
- DHCP..... 917

VRF Configuration..... 917
- Load VRF CAM..... 917
- Creating a Non-Default VRF Instance..... 917
- Assigning an Interface to a VRF..... 918
- View VRF Instance Information..... 918
- Assigning an OSPF Process to a VRF Instance.....918
- Configuring VRRP on a VRF Instance.....918

Sample VRF Configuration.....919
Route Leaking VRFs..... 927

55 Virtual Router Redundancy Protocol (VRRP).....928

VRRP Overview..... 928
VRRP Benefits.....929
VRRP Implementation..... 929
VRRP Configuration.....930
- Configuration Task List.....930
- Setting VRRP Initialization Delay.....940

Sample Configurations..... 941
- VRRP for an IPv4 Configuration.....941
- VRRP in a VRF Configuration.....945

56 S-Series Debugging and Diagnostics.....950

Offline Diagnostics..... 950
- Important Points to Remember..... 950
- Running Offline Diagnostics.....950

Configuring 104 10GbE Ports on the S6000 Platform.....955
- Configuring Fanout Mode..... 955
- Backward Compatibility With Enhanced Fanout Configuration.....956
- Configuring Fanout on Disabled Ports.....957

Trace Logs.....958
Auto Save on Crash or Rollover..... 958
Hardware Watchdog Timer.....959

Enabling Environmental Monitoring.......... 959
  Recognize an Overtemperature Condition.......... 959
  Troubleshoot an Over-temperature Condition.......... 960
  Recognize an Under-Voltage Condition.......... 960
  Troubleshoot an Under-Voltage Condition.......... 960
Buffer Tuning.......... 961
  Deciding to Tune Buffers.......... 963
  Using a Pre-Defined Buffer Profile.......... 966
  Sample Buffer Profile Configuration.......... 966
Troubleshooting Packet Loss.......... 967
  Displaying Drop Counters.......... 967
  Dataplane Statistics.......... 968
  Display Stack Port Statistics.......... 969
  Display Stack Member Counters.......... 970
Enabling Application Core Dumps.......... 970
Mini Core Dumps.......... 971
Enabling TCP Dumps.......... 972

57 Standards Compliance.......... 973
IEEE Compliance.......... 973
RFC and I-D Compliance.......... 974
  General Internet Protocols.......... 974
  General IPv4 Protocols.......... 975
  General IPv6 Protocols.......... 976
  Border Gateway Protocol (BGP).......... 976
  Open Shortest Path First (OSPF).......... 977
  Intermediate System to Intermediate System (IS-IS).......... 978
  Routing Information Protocol (RIP).......... 978
  Multicast.......... 979
  Network Management.......... 979
MIB Location.......... 986

1

About this Guide

This guide describes the protocols and features the Dell Networking Operating System (OS) supports and provides configuration instructions and examples for implementing them. This guide supports the S6000 platform.

The S6000 platform is available with Dell Networking OS version 9.0 (2.0) and beyond.

Though this guide contains information on protocols, it is not intended to be a complete reference. This guide is a reference for configuring protocols on Dell Networking systems. For complete information about protocols, refer to related documentation, including IETF requests for comments (RFCs). The instructions in this guide cite relevant RFCs. The Standards Compliance chapter contains a complete list of the supported RFCs and management information base files (MIBs).

Audience

This document is intended for system administrators who are responsible for configuring and maintaining networks and assumes knowledge in Layer 2 and Layer 3 networking technologies.

Conventions

This guide uses the following conventions to describe command syntax.

`Keyword`	Keywords are in Courier (a monospaced font) and must be entered in the CLI as listed.
parameter	Parameters are in italics and require a number or word to be entered in the CLI.
{X}	Keywords and parameters within braces must be entered in the CLI.
[X]	Keywords and parameters within brackets are optional.
x|y	Keywords and parameters separated by a bar require you to choose one option.
x||y	Keywords and parameters separated by a double bar allows you to choose any or all of the options.

Related Documents

For more information about the Dell Networking S6000 platform, refer the following documents:

- *Dell Networking OS Command Reference*
- *Installing the System*
- *Dell Quick Start Guide*
- *Dell Networking OS Release Notes*

2

Configuration Fundamentals

The Dell Networking Operating System (OS) command line interface (CLI) is a text-based interface you can use to configure interfaces and protocols.

The CLI is largely the same for the Z9000, S6000, S4810, and S4820T except for some commands and command outputs. The CLI is structured in modes for security and management purposes. Different sets of commands are available in each mode, and you can limit user access to modes using privilege levels.

In Dell Networking OS, after you enable a command, it is entered into the running configuration file. You can view the current configuration for the whole system or for a particular CLI mode. To save the current configuration, copy the running configuration to another location.

NOTE: Due to differences in hardware architecture and continued system development, features may occasionally differ between the platforms. Differences are noted in each CLI description and related documentation.

Accessing the Command Line

Access the CLI through a serial console port or a Telnet session.

When the system successfully boots, enter the command line in EXEC mode.

NOTE: You must have a password configured on a virtual terminal line before you can Telnet into the system. Therefore, you must use a console connection when connecting to the system for the first time.

```
telnet 172.31.1.53
Trying 172.31.1.53...
Connected to 172.31.1.53.
Escape character is '^]'.
Login: username
Password:
Dell>
```

CLI Modes

Different sets of commands are available in each mode.

A command found in one mode cannot be executed from another mode (except for EXEC mode commands with a preceding `do` command (refer to the `do` Command section).

You can set user access rights to commands and command modes using privilege levels;

The Dell Networking OS CLI is divided into three major mode levels:

- EXEC mode is the default mode and has a privilege level of 1, which is the most restricted level. Only a limited selection of commands is available, notably the `show` commands, which allow you to view system information.

- EXEC Privilege mode has commands to view configurations, clear counters, manage configuration files, run diagnostics, and enable or disable debug operations. The privilege level is 15, which is unrestricted. You can configure a password for this mode; refer to the *Configure the Enable Password* section in the Getting Started chapter.
- CONFIGURATION mode allows you to configure security features, time settings, set logging and SNMP functions, configure static ARP and MAC addresses, and set line cards on the system.

Beneath CONFIGURATION mode are submodes that apply to interfaces, protocols, and features. The following example shows the submode command structure. Two sub-CONFIGURATION modes are important when configuring the chassis for the first time:

- INTERFACE submode is the mode in which you configure Layer 2 and Layer 3 protocols and IP services specific to an interface. An interface can be physical (Management interface, 1 Gigabit Ethernet, or 10 Gigabit Ethernet, or synchronous optical network technologies [SONET]) or logical (Loopback, Null, port channel, or virtual local area network [VLAN]).
- LINE submode is the mode in which you to configure the console and virtual terminal lines.

NOTE: At any time, entering a question mark (`?`) displays the available command options. For example, when you are in CONFIGURATION mode, entering the question mark first lists all available commands, including the possible submodes.

The CLI modes are:

Navigating CLI Modes

The Dell Networking OS prompt changes to indicate the CLI mode.

The following table lists the CLI mode, its prompt, and information about how to access and exit the CLI mode. Move linearly through the command modes, except for the `end` command which takes you directly to EXEC Privilege mode and the `exit` command which moves you up one command mode level.

NOTE: Sub-CONFIGURATION modes all have the letters “conf” in the prompt with more modifiers to identify the mode and slot/port information.

Table 1. Dell Networking OS Command Modes

CLI Command Mode	Prompt	Access Command
EXEC	`Dell>`	Access the router through the console or Telnet.
EXEC Privilege	`Dell#`	• From EXEC mode, enter the `enable` command. • From any other mode, use the `end` command.
CONFIGURATION	`Dell(conf)#`	• From EXEC privilege mode, enter the `configure` command. • From every mode except EXEC and EXEC Privilege, enter the `exit` command.

CLI Command Mode	Prompt	Access Command
NOTE: Access all of the following modes from CONFIGURATION mode.		
AS-PATH ACL	`Dell(config-as-path)#`	`ip as-path access-list`
Gigabit Ethernet Interface	`Dell(conf-if-gi-0/0)#`	`interface` (INTERFACE modes)
10 Gigabit Ethernet Interface	`Dell(conf-if-te-0/1-2)#`	`interface` (INTERFACE modes)
Interface Group	`Dell(conf-if-group)#`	`interface(INTERFACE modes)`
Interface Range	`Dell(conf-if-range)#`	`interface` (INTERFACE modes)
Loopback Interface	`Dell(conf-if-lo-0)#`	`interface` (INTERFACE modes)
Management Ethernet Interface	`Dell(conf-if-ma-0/0)#`	`interface` (INTERFACE modes)
Null Interface	`Dell(conf-if-nu-0)#`	`interface` (INTERFACE modes)
Port-channel Interface	`Dell(conf-if-po-1)#`	`interface` (INTERFACE modes)
Tunnel Interface	`Dell(conf-if-tu-1)#`	`interface` (INTERFACE modes)
VLAN Interface	`Dell(conf-if-vl-1)#`	`interface` (INTERFACE modes)
STANDARD ACCESS-LIST	`Dell(config-std-nacl)#`	`ip access-list standard` (IP ACCESS-LIST Modes)
EXTENDED ACCESS-LIST	`Dell(config-ext-nacl)#`	`ip access-list extended` (IP ACCESS-LIST Modes)
IP COMMUNITY-LIST	`Dell(config-community-list)#`	`ip community-list`
AUXILIARY	`Dell(config-line-aux)#`	`line` (LINE Modes)
CONSOLE	`Dell(config-line-console)#`	`line` (LINE Modes)
VIRTUAL TERMINAL	`Dell(config-line-vty)#`	`line` (LINE Modes)
STANDARD ACCESS-LIST	`Dell(config-std-macl)#`	`mac access-list standard` (MAC ACCESS-LIST Modes)
EXTENDED ACCESS-LIST	`Dell(config-ext-macl)#`	`mac access-list extended` (MAC ACCESS-LIST Modes)
MULTIPLE SPANNING TREE	`Dell(config-mstp)#`	`protocol spanning-tree mstp`
Per-VLAN SPANNING TREE Plus	`Dell(config-pvst)#`	`protocol spanning-tree pvst`
PREFIX-LIST	`Dell(conf-nprefixl)#`	`ip prefix-list`

CLI Command Mode	Prompt	Access Command
RAPID SPANNING TREE	`Dell(config-rstp)#`	`protocol spanning-tree rstp`
REDIRECT	`Dell(conf-redirect-list)#`	`ip redirect-list`
ROUTE-MAP	`Dell(config-route-map)#`	`route-map`
ROUTER BGP	`Dell(conf-router_bgp)#`	`router bgp`
BGP ADDRESS-FAMILY	`Dell(conf-router_bgp_af)#` (for IPv4) `Dell(conf-routerZ_bgpv6_af)#` (for IPv6)	`address-family {ipv4 multicast | ipv6 unicast}` (ROUTER BGP Mode)
ROUTER ISIS	`Dell(conf-router_isis)#`	`router isis`
ISIS ADDRESS-FAMILY	`Dell(conf-router_isis-af_ipv6)#`	`address-family ipv6 unicast` (ROUTER ISIS Mode)
ROUTER OSPF	`Dell(conf-router_ospf)#`	`router ospf`
ROUTER OSPFV3	`Dell(conf-ipv6router_ospf)#`	`ipv6 router ospf`
ROUTER RIP	`Dell(conf-router_rip)#`	`router rip`
SPANNING TREE	`Dell(config-span)#`	`protocol spanning-tree 0`
TRACE-LIST	`Dell(conf-trace-acl)#`	`ip trace-list`
CLASS-MAP	`Dell(config-class-map)#`	`class-map`
CONTROL-PLANE	`Dell(conf-control-cpuqos)#`	`control-plane-cpuqos`
DCB POLICY	`Dell(conf-dcb-in)#` (for input policy) `Dell(conf-dcb-out)#` (for output policy)	`dcb-input` for input policy `dcb-output` for output policy
DHCP	`Dell(config-dhcp)#`	`ip dhcp server`
DHCP POOL	`Dell(config-dhcp-`*pool-name*`)#`	`pool` (DHCP Mode)
ECMP	`Dell(conf-ecmp-group-`*ecmp-group-id*`)#`	`ecmp-group`
EIS	`Dell(conf-mgmt-eis)#`	`management egress-interface-selection`
FRRP	`Dell(conf-frrp-`*ring-id*`)#`	`protocol frrp`
LLDP	`Dell(conf-lldp)#` or `Dell(conf-if—`*interface*`-lldp)#`	`protocol lldp` (CONFIGURATION or INTERFACE Modes)

CLI Command Mode	Prompt	Access Command
LLDP MANAGEMENT INTERFACE	`Dell(conf-lldp-mgmtIf)#`	`management-interface` (LLDP Mode)
LINE	`Dell(config-line-console)` or `Dell(config-line-vty)`	`line console` or `line vty`
MONITOR SESSION	`Dell(conf-mon-sess-`*`sessionID`*`)#`	`monitor session`
OPENFLOW INSTANCE	`Dell(conf-of-instance-`*`of-id`*`)#`	`openflow of-instance`
PORT-CHANNEL FAILOVER-GROUP	`Dell(conf-po-failover-grp)#`	`port-channel failover-group`
PRIORITY GROUP	`Dell(conf-pg)#`	`priority-group`
PROTOCOL GVRP	`Dell(config-gvrp)#`	`protocol gvrp`
QOS POLICY	`Dell(conf-qos-policy-out-ets)#`	`qos-policy-output`
VLT DOMAIN	`Dell(conf-vlt-domain)#`	`vlt domain`
VRRP	`Dell(conf-if-`*`interface-type-slot/port`*`-vrid-`*`vrrp-group-id`*`)#`	`vrrp-group`
u-Boot	`Dell=>`	Press any key when the following line appears on the console during a system boot: `Hit any key to stop autoboot:`
UPLINK STATE GROUP	`Dell(conf-uplink-state-group-`*`groupID`*`)#`	`uplink-state-group`

The following example shows how to change the command mode from CONFIGURATION mode to PROTOCOL SPANNING TREE.

Example of Changing Command Modes

```
Dell(conf)#protocol spanning-tree 0
Dell(config-span)#
```

The do Command

You can enter an EXEC mode command from any CONFIGURATION mode (CONFIGURATION, INTERFACE, SPANNING TREE, and so on.) without having to return to EXEC mode by preceding the EXEC mode command with the `do` command.

The following example shows the output of the `do` command.

```
Dell(conf)#do show system brief

Stack MAC : 00:01:e8:00:66:64
Reload-Type                :   normal-reload [Next boot : normal-reload]
```

```
--  Stack Info  --
Unit  UnitType     Status         ReqTyp          CurTyp          Version
Ports
--------------------------------------------------------------------------------
----
  0   Management   online         S4810           S4810           9.4(0.0)    64
  1   Member       not present
  2   Member       not present
  3   Member       not present
  4   Member       not present
  5   Member       not present
  6   Member       not present
  7   Member       not present
  8   Member       not present
  9   Member       not present
 10   Member       not present
 11   Member       not present

--  Power Supplies  --
Unit   Bay   Status       Type    FanStatus
---------------------------------------------------------------------------
  0     0    absent               absent
  0     1    up           UNKNOWN up

--  Fan  Status  --
Unit Bay   TrayStatus  Fan0    Speed   Fan1    Speed
--------------------------------------------------------------------------------
----
 0    0     up          up      9120    up      9120
 0    1     up          up      9120    up      9120

Speed in RPM

Dell(conf)#
```

Undoing Commands

When you enter a command, the command line is added to the running configuration file (running-config).

To disable a command and remove it from the running-config, enter the `no` command, then the original command. For example, to delete an IP address configured on an interface, use the `no ip address` *`ip-address`* command.

NOTE: Use the help or `?` command as described in Obtaining Help.

Example of Viewing Disabled Commands

```
Dell(conf)#interface tengigabitethernet 4/17
Dell(conf-if-te-4/17)#ip address 192.168.10.1/24
Dell(conf-if-te-4/17)#show config
!
  interface tenGigabitEthernet 4/17
  ip address 192.168.10.1/24
no shutdown
Dell(conf-if-te-4/17)#no ip address
Dell(conf-if-te-4/17)#show config
!
interface tenGigabitEthernet 4/17
```

```
no ip address
no shutdown
```

Layer 2 protocols are disabled by default. To enable Layer 2 protocols, use the `no disable` command. For example, in PROTOCOL SPANNING TREE mode, enter `no disable` to enable Spanning Tree.

Obtaining Help

Obtain a list of keywords and a brief functional description of those keywords at any CLI mode using the `?` or help command:

- To list the keywords available in the current mode, enter `?` at the prompt or after a keyword.
- Enter `?` after a prompt lists all of the available keywords. The output of this command is the same for the help command.

```
Dell#?
cd                              Change current directory
clear                           Reset functions
clock                           Manage the system clock
configure                       Configuring from terminal
copy                            Copy from one file to another
debug                           Debug functions
--More--
```

- Enter `?` after a partial keyword lists all of the keywords that begin with the specified letters.

```
Dell(conf)#cl?
class-map
clock
Dell(conf)#cl
```

- Enter `[space]?` after a keyword lists all of the keywords that can follow the specified keyword.

```
Dell(conf)#clock ?
summer-time       Configure summer (daylight savings) time
timezone          Configure time zone
Dell(conf)#clock
```

Entering and Editing Commands

Notes for entering commands.

- The CLI is not case-sensitive.
- You can enter partial CLI keywords.
  - Enter the minimum number of letters to uniquely identify a command. For example, you cannot enter `cl` as a partial keyword because both the `clock` and `class-map` commands begin with the letters “cl.” You can enter `clo`, however, as a partial keyword because only one command begins with those three letters.
- The TAB key auto-completes keywords in commands. Enter the minimum number of letters to uniquely identify a command.
- The UP and DOWN arrow keys display previously entered commands (refer to Command History).
- The BACKSPACE and DELETE keys erase the previous letter.
- Key combinations are available to move quickly across the command line. The following table describes these short-cut key combinations.

Short-Cut Key Combination	Action
CNTL-A	Moves the cursor to the beginning of the command line.
CNTL-B	Moves the cursor back one character.
CNTL-D	Deletes character at cursor.
CNTL-E	Moves the cursor to the end of the line.
CNTL-F	Moves the cursor forward one character.
CNTL-I	Completes a keyword.
CNTL-K	Deletes all characters from the cursor to the end of the command line.
CNTL-L	Re-enters the previous command.
CNTL-N	Return to more recent commands in the history buffer after recalling commands with CTRL-P or the UP arrow key.
CNTL-P	Recalls commands, beginning with the last command.
CNTL-R	Re-enters the previous command.
CNTL-U	Deletes the line.
CNTL-W	Deletes the previous word.
CNTL-X	Deletes the line.
CNTL-Z	Ends continuous scrolling of command outputs.
Esc B	Moves the cursor back one word.
Esc F	Moves the cursor forward one word.
Esc D	Deletes all characters from the cursor to the end of the word.

Command History

Dell Networking OS maintains a history of previously-entered commands for each mode. For example:

- When you are in EXEC mode, the UP and DOWN arrow keys display the previously-entered EXEC mode commands.
- When you are in CONFIGURATION mode, the UP or DOWN arrows keys recall the previously-entered CONFIGURATION mode commands.

Filtering show Command Outputs

Filter the output of a `show` command to display specific information by adding `| [except | find | grep | no-more | save]` *`specified_text`* after the command.

The variable *`specified_text`* is the text for which you are filtering and it IS case sensitive unless you use the `ignore-case` sub-option.

Starting with Dell Networking OS version 7.8.1.0, the `grep` command accepts an `ignore-case` sub-option that forces the search to case-insensitive. For example, the commands:

- `show run | grep Ethernet` returns a search result with instances containing a capitalized "Ethernet," such as `interface GigabitEthernet 0/0`.
- `show run | grep ethernet` does not return that search result because it only searches for instances containing a non-capitalized "ethernet."
- `show run | grep Ethernet ignore-case` returns instances containing both "Ethernet" and "ethernet."

The `grep` command displays only the lines containing specified text. The following example shows this command used in combination with the `show linecard all` command.

```
Dell(conf)#do show system brief | grep 0
0     not present
```

NOTE: Dell Networking OS accepts a space or no space before and after the pipe. To filter a phrase with spaces, underscores, or ranges, enclose the phrase with double quotation marks.

The `except` keyword displays text that does not match the specified text. The following example shows this command used in combination with the `show linecard all` command.

Example of the `except` Keyword

```
Dell#show system brief | except 0

Slot Status NxtBoot ReqTyp CurTyp Version Ports
-----------------------------------------------------
2    not present
3    not present
4    not present
5    not present
6    not present
```

The `find` keyword displays the output of the `show` command beginning from the first occurrence of specified text. The following example shows this command used in combination with the `show linecard all` command.

Example of the `find` Keyword

```
Dell(conf)#do show system brief | find 0
0  not present
1  not present
2  online   online  E48TB  E48TB  1-1-463  48
3  not present
4  not present
5  online   online  E48VB  E48VB  1-1-463  48
6  not present
7  not present
```

The `display` command displays additional configuration information.

The `no-more` command displays the output all at once rather than one screen at a time. This is similar to the `terminal length` command except that the `no-more` option affects the output of the specified command only.

The `save` command copies the output to a file for future reference.

NOTE: You can filter a single command output multiple times. The `save` option must be the last option entered. For example: `Dell# command | grep regular-expression | except regular-expression | grep other-regular-expression | find regular-expression | save`.

Multiple Users in Configuration Mode

Dell Networking OS notifies all users when there are multiple users logged in to CONFIGURATION mode.

A warning message indicates the username, type of connection (console or VTY), and in the case of a VTY connection, the IP address of the terminal on which the connection was established. For example:

- On the system that telnets into the switch, this message appears:

  ```
  % Warning: The following users are currently configuring the system:
  User "<username>" on line console0
  ```

- On the system that is connected over the console, this message appears:

  ```
  % Warning: User "<username>" on line vty0 "10.11.130.2" is in configuration
  mode
  ```

If either of these messages appears, Dell Networking recommends coordinating with the users listed in the message so that you do not unintentionally overwrite each other’s configuration changes.

3

Getting Started

This chapter describes how you start configuring your system.

When you power up the chassis, the system performs a power-on self test (POST) during which the line card status light emitting diodes (LEDs) blink green. The system then loads the Dell Networking Operating System (OS). Boot messages scroll up the terminal window during this process. No user interaction is required if the boot process proceeds without interruption.

When the boot process completes, the RPM and line card status LEDs remain online (green) and the console monitor displays the EXEC mode prompt.

For details about using the command line interface (CLI), refer to the Accessing the Command Line section in the Configuration Fundamentals chapter.

Console Access

The has two management ports available for system access: a serial console port and an out-of-bounds (OOB) port.

Serial Console

The RJ-45/RS-232 console port is labeled on the chassis. It is in the upper right-hand side, as you face the I/O side of the chassis.

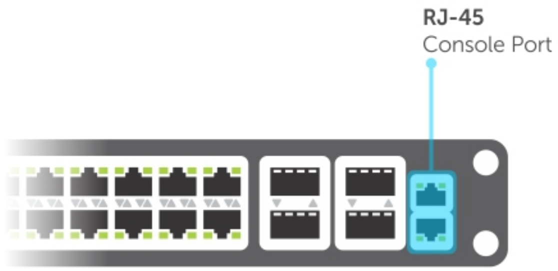


Figure 1. RJ-45 Console Port

1. RJ-45 Console Port

Accessing the Console Port

To access the console port, follow these steps:
For the console port pinout, refer to Accessing the RJ-45 Console Port with a DB-9 Adapter.

1. Install an RJ-45 copper cable into the console port.Use a rollover (crossover) cable to connect the S4810 console port to a terminal server.
2. Connect the other end of the cable to the DTE terminal server.
3. Terminal settings on the console port cannot be changed in the software and are set as follows:
    - 9600 baud rate
    - No parity
    - 8 data bits
    - 1 stop bit
    - No flow control

Pin Assignments

You can connect to the console using a RJ-45 to RJ-45 rollover cable and a RJ-45 to DB-9 female DTE adapter to a terminal server (for example, a PC).

The pin assignments between the console and a DTE terminal server are as follows:

Table 2. Pin Assignments Between the Console and a DTE Terminal Server

Console Port	RJ-45 to RJ-45 Rollover Cable	RJ-45 to RJ-45 Rollover Cable	RJ-45 to DB-9 Adapter	Terminal Server Device
Signal	**RJ-45 Pinout**	**RJ-45 Pinout**	**DB-9 Pin**	**Signal**
RTS	1	8	8	CTS
NC	2	7	6	DSR
TxD	3	6	2	RxD
GND	4	5	5	GND
GND	5	4	5	GND
RxD	6	3	3	TxD
NC	7	2	4	DTR
CTS	8	1	7	RTS

Default Configuration

A version of Dell Networking OS is pre-loaded onto the chassis; however, the system is not configured when you power up for the first time (except for the default hostname, which is `Dell`). You must configure the system using the CLI.

Configuring a Host Name

The host name appears in the prompt. The default host name is `Dell`.

- Host names must start with a letter and end with a letter or digit.

- Characters within the string can be letters, digits, and hyphens.

To create a host name, use the following command.

- Create a host name.

  CONFIGURATION mode

  ```
  hostname name
  ```

Example of the `hostname` Command

```
Dell(conf)#hostname R1
R1(conf)#
```

Accessing the System Remotely

You can configure the system to access it remotely by Telnet or SSH.

- The S6000has a dedicated management port and a management routing table that is separate from the IP routing table.
- You can manage all Dell Networking products in-band via the front-end data ports through interfaces assigned an IP address as well.

Accessing the and S6000Remotely

Configuring the system for Telnet is a three-step process, as described in the following topics:

1. Configure an IP address for the management port. Configure the Management Port IP Address
2. Configure a management route with a default gateway. Configure a Management Route
3. Configure a username and password. Configure a Username and Password

Configure the Management Port IP Address

To access the system remotely, assign IP addresses to the management ports.

1. Enter INTERFACE mode for the Management port.

   CONFIGURATION mode

   ```
   interface ManagementEthernet slot/port
   ```

   - *`slot`*: the range is from 0 to 11.
   - *`port`*: the range is 0.
2. Assign an IP address to the interface.

   INTERFACE mode

   ```
   ip address ip-address/mask
   ```

   - *`ip-address`*: an address in dotted-decimal format (A.B.C.D).
   - *`mask`*: a subnet mask in /prefix-length format (/ xx).
3. Enable the interface.

   INTERFACE mode

   ```
   no shutdown
   ```

Configure a Management Route

Define a path from the system to the network from which you are accessing the system remotely. Management routes are separate from IP routes and are only used to manage the system through the management port.
To configure a management route, use the following command.

- Configure a management route to the network from which you are accessing the system.
  CONFIGURATION mode

  ```
  management route ip-address/mask gateway
  ```

  - `ip-address`: the network address in dotted-decimal format (A.B.C.D).
  - `mask`: a subnet mask in /prefix-length format (/ xx).
  - `gateway`: the next hop for network traffic originating from the management port.

Configuring a Username and Password

To access the system remotely, configure a system username and password.
To configure a system username and password, use the following command.

- Configure a username and password to access the system remotely.
  CONFIGURATION mode

  ```
  username username password [encryption-type] password
  ```

  - `encryption-type`: specifies how you are inputting the password, is 0 by default, and is not required.
    - 0 is for inputting the password in clear text.
    - 7 is for inputting a password that is already encrypted using a Type 7 hash. Obtaining the encrypted password from the configuration of another Dell Networking system.

Configuring the Enable Password

Access EXEC Privilege mode using the `enable` command. EXEC Privilege mode is unrestricted by default. Configure a password as a basic security measure.

There are two types of enable passwords:

- `enable password` stores the password in the running/startup configuration using a DES encryption method.
- `enable secret` is stored in the running/startup configuration in using a stronger, MD5 encryption method.

Dell Networking recommends using the `enable secret` password.

To configure an enable password, use the following command.

- Create a password to access EXEC Privilege mode.
  CONFIGURATION mode

  ```
  enable [password | secret] [level level] [encryption-type] password
  ```

- *`level`*: is the privilege level, is 15 by default, and is not required
- *`encryption-type`*: specifies how you are inputting the password, is 0 by default, and is not required.
  - 0 is for inputting the password in clear text.
  - 7 is for inputting a password that is already encrypted using a DES hash. Obtain the encrypted password from the configuration file of another Dell Networking system.
  - 5 is for inputting a password that is already encrypted using an MD5 hash. Obtain the encrypted password from the configuration file of another Dell Networking system.

Configuration File Management

Files can be stored on and accessed from various storage media. Rename, delete, and copy files on the system from EXEC Privilege mode.

Copy Files to and from the System

The command syntax for copying files is similar to UNIX. The copy command uses the format `copy` *`source-file-url destination-file-url`*.

NOTE: For a detailed description of the copy command, refer to the *Dell Networking OS Command Reference*.

- To copy a local file to a remote system, combine the file-origin syntax for a local file location with the file-destination syntax for a remote file location.
- To copy a remote file to Dell Networking system, combine the file-origin syntax for a remote file location with the file-destination syntax for a local file location.

Table 3. Forming a `copy` Command

Location	*source-file-url* Syntax	*destination-file-url* Syntax
For a remote file location: FTP server	`copy ftp:// username:password@{hostip | hostname}/filepath/ filename`	`ftp:// username:password@{hostip | hostname}/ filepath/ filename`
For a remote file location: TFTP server	`copy tftp://{hostip | hostname}/filepath/ filename`	`tftp://{hostip | hostname}/filepath/ filename`
For a remote file location: SCP server	`copy scp://{hostip | hostname}/filepath/ filename`	`scp://{hostip | hostname}/filepath/ filename`

Important Points to Remember

- You may not copy a file from one remote system to another.
- You may not copy a file from one location to the same location.
- When copying to a server, you can only use a hostname if a domain name server (DNS) server is configured.
- The `usbflash` command is supported on S6000. Refer to your system's Release Notes for a list of approved USB vendors.

Example of Copying a File to an FTP Server

```
Dell#copy flash://Dell-EF-8.2.1.0.bin ftp://myusername:mypassword@10.10.10.10/
/Dell/Dell-EF-8.2.1.0
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
27952672 bytes successfully copied
```

Example of Importing a File to the Local System

```
core1#$//copy ftp://myusername:mypassword@10.10.10.10//Dell/
Dell-EF-8.2.1.0.bin flash://
Destination file name [Dell-EF-8.2.1.0.bin.bin]:
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
26292881 bytes successfully copied
```

Save the Running-Configuration

The running-configuration contains the current system configuration. Dell Networking recommends coping your running-configuration to the startup-configuration.
The commands in this section follow the same format as those commands in the Copy Files to and from the System section but use the filenames startup-configuration and running-configuration. These commands assume that current directory is the internal flash, which is the system default.

- Save the running-configuration to the startup-configuration on the internal flash of the primary RPM.

  EXEC Privilege mode

  `copy running-config startup-config`
- Save the running-configuration to the internal flash on an RPM.

  EXEC Privilege mode

  `copy running-config rpm{0|1}flash://`*`filename`*
- Save the running-configuration to an FTP server.

  EXEC Privilege mode

  `copy running-config ftp://` *`username:password@{hostip | hostname}/filepath/filename`*
- Save the running-configuration to a TFTP server.

  EXEC Privilege mode

  `copy running-config tftp://`*`{hostip | hostname}/ filepath/filename`*
- Save the running-configuration to an SCP server.

  EXEC Privilege mode

  `copy running-config scp://`*`{hostip | hostname}/ filepath/filename`*

  NOTE: When copying to a server, a host name can only be used if a DNS server is configured.

Configure the Overload Bit for a Startup Scenario

For information about setting the router overload bit for a specific period of time after a switch reload is implemented, refer to the *Intermediate System to Intermediate System (IS-IS)* section in the *Dell Networking OS Command Line Reference Guide*.

Viewing Files

You can only view file information and content on local file systems.
To view a list of files or the contents of a file, use the following commands.

- View a list of files on the internal flash.

  EXEC Privilege mode

  ```
  dir flash:
  ```
- View the running-configuration.

  EXEC Privilege mode

  ```
  show running-config
  ```
- View the startup-configuration.

  EXEC Privilege mode

  ```
  show startup-config
  ```

Example of the `dir` Command

The output of the `dir` command also shows the read/write privileges, size (in bytes), and date of modification for each file.

```
Dell#dir
Directory of flash:

 1 drw-    32768 Jan 01 1980 00:00:00 .
 2 drwx      512 Jul 23 2007 00:38:44 ..
 3 drw-     8192 Mar 30 1919 10:31:04 TRACE_LOG_DIR
 4 drw-     8192 Mar 30 1919 10:31:04 CRASH_LOG_DIR
 5 drw-     8192 Mar 30 1919 10:31:04 NVTRACE_LOG_DIR
 6 drw-     8192 Mar 30 1919 10:31:04 CORE_DUMP_DIR
 7 d---     8192 Mar 30 1919 10:31:04 ADMIN_DIR
 8 -rw- 33059550 Jul 11 2007 17:49:46 FTOS-EF-7.4.2.0.bin
 9 -rw- 27674906 Jul 06 2007 00:20:24 FTOS-EF-4.7.4.302.bin
10 -rw- 27674906 Jul 06 2007 19:54:52 boot-image-FILE
11 drw-     8192 Jan 01 1980 00:18:28 diag
12 -rw-     7276 Jul 20 2007 01:52:40 startup-config.bak
13 -rw-     7341 Jul 20 2007 15:34:46 startup-config
14 -rw- 27674906 Jul 06 2007 19:52:22 boot-image
15 -rw- 27674906 Jul 06 2007 02:23:22 boot-flash
--More--
```

View Configuration Files

Configuration files have three commented lines at the beginning of the file, as shown in the following example, to help you track the last time any user made a change to the file, which user made the changes, and when the file was last saved to the startup-configuration.

In the running-configuration file, if there is a difference between the timestamp on the "Last configuration change" and "Startup-config last updated," you have made changes that have not been saved and are preserved after a system reboot.

Example of the `show running-config` Command

```
Dell#show running-config
Current Configuration ...
! Version 9.4(0.0)
! Last configuration change at Tue Mar 11 21:33:56 2014 by admin
! Startup-config last updated at Tue Mar 11 12:11:00 2014 by default
!
boot system stack-unit 0 primary system: B:
boot system stack-unit 0 secondary tftp://10.16.127.35/dt-maa-s4810-2
boot system stack-unit 0 default tftp://10.16.127.35/dt-maa-s4810-2
boot system gateway 10.16.130.254
!

Page 57    - Under Managing the File System, the word external Flash must be
removed

Page 57    - The output of show file-systems must be modified as follows.

Dell#show file-systems

      Size(b)     Free(b)      Feature      Type   Flags  Prefixes
   2056916992  2056540160        FAT32 USERFLASH      rw  flash:
            -           -            -   network      rw  ftp:
            -           -            -   network      rw  tftp:
            -           -            -   network      rw  scp:
Dell#
```

Managing the File System

The Dell Networking system can use the internal Flash, external Flash, or remote devices to store files. The system stores files on the internal Flash by default but can be configured to store files elsewhere.

To view file system information, use the following command.

- View information about each file system.
  EXEC Privilege mode

  `show file-systems`

The output of the `show file-systems` command in the following example shows the total capacity, amount of free memory, file structure, media type, read/write privileges for each storage device in use.

```
Dell#show file-systems
Size(b)   Free(b)   Feature Type Flags Prefixes
520962048 213778432 dosFs2.0 USERFLASH rw flash:
127772672  21936128 dosFs2.0 USERFLASH rw slot0:
        -         -        - network   rw ftp:
```

```
        -         -        - network   rw tftp:
        -         -        - network   rw scp:
```

You can change the default file system so that file management commands apply to a particular device or memory.

To change the default directory, use the following command.

- Change the default directory.
  EXEC Privilege mode

  `cd directory`

View Command History

The command-history trace feature captures all commands entered by all users of the system with a time stamp and writes these messages to a dedicated trace log buffer.

The system generates a trace message for each executed command. No password information is saved to the file.

To view the command-history trace, use the `show command-history` command.

Example of the `show command-history` Command

```
Dell#show command-history
[12/5 10:57:8]: CMD-(CLI):service password-encryption
[12/5 10:57:12]: CMD-(CLI):hostname Force10
[12/5 10:57:12]: CMD-(CLI):ip telnet server enable
[12/5 10:57:12]: CMD-(CLI):line console 0
[12/5 10:57:12]: CMD-(CLI):line vty 0 9
[12/5 10:57:13]: CMD-(CLI):boot system rpm0 primary flash://FTOS-
CB-1.1.1.2E2.bin
```

Upgrading Dell Networking OS

NOTE: To upgrade Dell Networking Operating System (OS), refer to the Release Notes for the version you want to load on the system.

Using HTTP for File Transfers

Stating with Release 9.3(0.1), you can use HTTP to copy files or configuration details to a remote server. Use the **copy source-file-url http://host[:port]/file-path** command to transfer files to an external server. This functionality to transport files using HTTP to a remote server is supported on MXL, I/O Aggregator, S4810, S4820, S6000, and Z9000 platforms.
Enter the following *source-file-url* keywords and information:

- To copy a file from the internal FLASH, enter flash:// followed by the filename.
- To copy the running configuration, enter the keyword running-config.
- To copy the startup configuration, enter the keyword startup-config.
- To copy a file on the external FLASH, enter usbflash:// followed by the filename.

Using Hashes to Validate Software Images

You can use the MD5 message-digest algorithm or SHA256 Secure Hash Algorithm to validate the software image on the flash drive, after the image has been transferred to the system, but before the image has been installed. The validation calculates a hash value of the downloaded image file on system's flash drive, and**,** optionally**,** compares it to a Dell Networking published hash for that file.

The MD5 or SHA256 hash provides a method of validating that you have downloaded the original software. Calculating the hash on the local image file, and comparing the result to the hash published for that file on iSupport, provides a high level of confidence that the local copy is exactly the same as the published software image. This validation procedure, and the **verify** {**md5** | **sha256**} command to support it, can prevent the installation of corrupted or modified images.

The **verify** {**md5** | **sha256**} command calculates and displays the hash of any file on the specified local flash drive. You can compare the displayed hash against the appropriate hash published on i-Support. Optionally, the published hash can be included in the **verify** {**md5** | **sha256**} command, which will display whether it matches the calculated hash of the indicated file.

To validate a software image:

1. Download Dell Networking OS software image file from the iSupport page to the local (FTP or TFTP) server. The published hash for that file is displayed next to the software image file on the iSupport page.
2. Go on to the Dell Networking system and copy the software image to the flash drive, using the **copy** command.
3. Run the **verify** {**md5** | **sha256**} [**flash**://]*img-file* [*hash-value*] command. For example, **verify sha256 flash://FTOS-SE-9.5.0.0.bin**
4. Compare the generated hash value to the expected hash value published on the iSupport page.

To validate the software image on the flash drive after the image has been transferred to the system, but before the image has been installed, use the **verify** {**md5** | **sha256**} [**flash**://]*img-file* [*hash-value*] command in EXEC mode.

- **md5**: MD5 message-digest algorithm
- **sha256**: SHA256 Secure Hash Algorithm
- **flash:** (Optional) Specifies the flash drive. The default is to use the flash drive. You can just enter the image file name.
- *hash-value*: (Optional). Specify the relevant hash published on i-Support.
- *img-file*: Enter the name **of** the Dell Networking **software** image file to validate

Examples: Without Entering the Hash Value for Verification

MD5

```
Dell# verify md5 flash://FTOS-SE-9.5.0.0.bin
MD5 hash for FTOS-SE-9.5.0.0.bin: 275ceb73a4f3118e1d6bcf7d75753459
```

SHA256

```
Dell# verify sha256 flash://FTOS-SE-9.5.0.0.bin
SHA256 hash for FTOS-SE-9.5.0.0.bin:
e6328c06faf814e6899ceead219afbf9360e986d692988023b749e6b2093e933
```

Examples: Entering the Hash Value for Verification

MD5

```
Dell# verify md5 flash://FTOS-SE-9.5.0.0.bin 275ceb73a4f3118e1d6bcf7d75753459
MD5 hash VERIFIED for FTOS-SE-9.5.0.0.bin
```

SHA256

```
Dell# verify sha256 flash://FTOS-SE-9.5.0.0.bin
e6328c06faf814e6899ceead219afbf9360e986d692988023b749e6b2093e933
SHA256 hash VERIFIED for FTOS-SE-9.5.0.0.bin
```

4

Management

Management is supported on the S6000 platform.

This chapter describes the different protocols or services used to manage the Dell Networking system.

Configuring Privilege Levels

Privilege levels restrict access to commands based on user or terminal line.

There are 16 privilege levels, of which three are pre-defined. The default privilege level is **1**.

Level	Description
Level 0	Access to the system begins at EXEC mode, and EXEC mode commands are limited to `enable`, `disable`, and `exit`.
Level 1	Access to the system begins at EXEC mode, and all commands are available.
Level 15	Access to the system begins at EXEC Privilege mode, and all commands are available.

Creating a Custom Privilege Level

Custom privilege levels start with the default EXEC mode command set. You can then customize privilege levels 2-14 by:

- restricting access to an EXEC mode command
- moving commands from EXEC Privilege to EXEC mode
- restricting access

A user can access all commands at his privilege level and below.

Removing a Command from EXEC Mode

To remove a command from the list of available commands in EXEC mode for a specific privilege level, use the `privilege exec` command from CONFIGURATION mode.

In the command, specify a level *greater* than the level given to a user or terminal line, then the first keyword of each command you wish to restrict.

Moving a Command from EXEC Privilege Mode to EXEC Mode

To move a command from EXEC Privilege to EXEC mode for a privilege level, use the `privilege exec` command from CONFIGURATION mode.

In the command, specify the privilege level of the user or terminal line and specify *all* keywords in the command to which you want to allow access.

Allowing Access to CONFIGURATION Mode Commands

To allow access to CONFIGURATION mode, use the `privilege exec level level configure` command from CONFIGURATION mode.

A user that enters CONFIGURATION mode remains at his privilege level and has access to only two commands, `end` and `exit`. You must individually specify each CONFIGURATION mode command you want to allow access to using the `privilege configure level level` command. In the command, specify the privilege level of the user or terminal line and specify *all* the keywords in the command to which you want to allow access.

Allowing Access to the Following Modes

This section describes how to allow access to the INTERFACE, LINE, ROUTE-MAP, and ROUTER modes. Similar to allowing access to CONFIGURATION mode, to allow access to INTERFACE, LINE, ROUTE-MAP, and ROUTER modes, you must first allow access to the command that enters you into the mode. For example, to allow a user to enter INTERFACE mode, use the `privilege configure level level interface gigabitethernet` command.

Next, individually identify the INTERFACE, LINE, ROUTE-MAP or ROUTER commands to which you want to allow access using the `privilege {interface | line | route-map | router} level level` command. In the command, specify the privilege level of the user or terminal line and specify *all* the keywords in the command to which you want to allow access.

To remove, move or allow access, use the following commands.

The configuration in the following example creates privilege level 3. This level:

- removes the `resequence` command from EXEC mode by requiring a minimum of privilege level 4
- moves the `capture bgp-pdu max-buffer-size` command from EXEC Privilege to EXEC mode by requiring a minimum privilege level 3, which is the configured level for VTY 0
- allows access to CONFIGURATION mode with the `banner` command
- allows access to INTERFACE and LINE modes are allowed with no commands

- Remove a command from the list of available commands in EXEC mode.

  CONFIGURATION mode

  ```
  privilege exec level level {command ||...|| command}
  ```
- Move a command from EXEC Privilege to EXEC mode.

  CONFIGURATION mode

  ```
  privilege exec level level {command ||...|| command}
  ```
- Allow access to CONFIGURATION mode.

  CONFIGURATION mode

  ```
  privilege exec level level configure
  ```
- Allow access to INTERFACE, LINE, ROUTE-MAP, and/or ROUTER mode. Specify *all* the keywords in the command.

  CONFIGURATION mode

  ```
  privilege configure level level {interface | line | route-map | router}
  {command-keyword ||...|| command-keyword}
  ```

- Allow access to a CONFIGURATION, INTERFACE, LINE, ROUTE-MAP, and/or ROUTER mode command.

  CONFIGURATION mode

  `privilege {configure |interface | line | route-map | router} level` *`level`* `{`*`command`* `||...||` *`command`*`}`

Example of EXEC Privilege Commands

```
Dell(conf)#do show run priv
!
privilege exec level 3 capture
privilege exec level 3 configure
privilege exec level 4 resequence
privilege exec level 3 capture bgp-pdu
privilege exec level 3 capture bgp-pdu max-buffer-size
privilege configure level 3 line
privilege configure level 3 interface
Dell(conf)#do telnet 10.11.80.201
[telnet output omitted]
Dell#show priv
Current privilege level is 3.
Dell#?
capture         Capture packet
configure       Configuring from terminal
disable         Turn off privileged commands
enable          Turn on privileged commands
exit            Exit from the EXEC
ip              Global IP subcommands
monitor         Monitoring feature
mtrace          Trace reverse multicast path from destination to source
ping            Send echo messages
quit            Exit from the EXEC
show            Show running system information
[output omitted]
Dell#config
[output omitted]
Dell(conf)#do show priv
Current privilege level is 3.
Dell(conf)#?
end                 Exit from configuration mode
exit                Exit from configuration mode
interface           Select an interface to configure
line                Configure a terminal line
linecard            Set line card type
Dell(conf)#interface ?
fastethernet        Fast Ethernet interface
gigabitethernet     Gigabit Ethernet interface
loopback            Loopback interface
managementethernet  Management Ethernet interface
null                Null interface
port-channel        Port-channel interface
range               Configure interface range
sonet               SONET interface
tengigabitethernet  TenGigabit Ethernet interface
vlan                VLAN interface
Dell(conf)#interface gigabitethernet 1/1
Dell(conf-if-gi-1/1)#?
end                 Exit from configuration mode
exit                Exit from interface configuration mode
Dell(conf-if-gi-1/1)#exit
Dell(conf)#line ?
```

```
aux                 Auxiliary line
console             Primary terminal line
vty                 Virtual terminal
Dell(conf)#line vty 0
Dell(config-line-vty)#?
exit                Exit from line configuration mode
Dell(config-line-vty)#
Dell(conf)#interface group ?
fortyGigE                 FortyGigabit Ethernet interface
gigabitethernet         GigabitEthernet interface IEEE 802.3z
tengigabitethernet     TenGigabit Ethernet interface
vlan                         VLAN keyword
Dell(conf)# interface group vlan 1 - 2 , tengigabitethernet 0/0
Dell(conf-if-group-vl-1-2,te-0/0)# no shutdown
Dell(conf-if-group-vl-1-2,te-0/0)# end
```

Applying a Privilege Level to a Username

To set the user privilege level, use the following command.

- Configure a privilege level for a user.

  CONFIGURATION mode

  `username *username* privilege *level*`

Applying a Privilege Level to a Terminal Line

To set a privilege level for a terminal line, use the following command.

- Configure a privilege level for a user.

  CONFIGURATION mode

  `username username privilege *level*`

NOTE: When you assign a privilege level between 2 and 15, access to the system begins at EXEC mode, but the prompt is `hostname#`, rather than `hostname>`.

Configuring Logging

The Dell Networking OS tracks changes in the system using event and error messages.
By default, Dell Networking OS logs these messages on:

- the internal buffer
- console and terminal lines
- any configured syslog servers

To disable logging, use the following commands.

- Disable all logging except on the console.

  CONFIGURATION mode

  `no logging on`
- Disable logging to the logging buffer.

  CONFIGURATION mode

  `no logging buffer`

- Disable logging to terminal lines.

  CONFIGURATION mode

  ```
  no logging monitor
  ```
- Disable console logging.

  CONFIGURATION mode

  ```
  no logging console
  ```

Audit and Security Logs

This section describes how to configure, display, and clear audit and security logs.

The following is the configuration task list for audit and security logs:

- Enabling Audit and Security Logs
- Displaying Audit and Security Logs
- Clearing Audit Logs

Enabling Audit and Security Logs

You enable audit and security logs to monitor configuration changes or determine if these changes affect the operation of the system in the network. You log audit and security events to a system log server, using the **logging extended** command in CONFIGURATION mode.

Audit Logs

The audit log contains configuration events and information. The types of information in this log consist of the following:

- User logins to the switch.
- System events for network issues or system issues.
- Users making configuration changes. The switch logs who made the configuration changes and the date and time of the change. However, each specific change on the configuration is not logged. Only that the configuration was modified is logged with the user ID, date, and time of the change.
- Uncontrolled shutdown.

Security Logs

The security log contains security events and information. RBAC restricts access to audit and security logs based on the CLI sessions' user roles. The types of information in this log consist of the following:

- Establishment of secure traffic flows, such as SSH.
- Violations on secure flows or certificate issues.
- Adding and deleting of users.
- User access and configuration changes to the security and crypto parameters (not the key information but the crypto configuration)

Important Points to Remember

When you enabled RBAC and extended logging:

- Only the system administrator user role can execute this command.
- The system administrator and system security administrator user roles can view security events and system events.
- The system administrator user roles can view audit, security, and system events.
- Only the system administrator and security administrator user roles can view security logs.
- The network administrator and network operator user roles can view system events.

NOTE: If extended logging is disabled, you can only view system events, regardless of RBAC user role.

Example of Enabling Audit and Security Logs

```
Dell(conf)#logging extended
```

Displaying Audit and Security Logs

To display audit logs, use the `show logging auditlog` command in Exec mode. To view these logs, you must first enable the logging extended command. Only the RBAC system administrator user role can view the audit logs. Only the RBAC security administrator and system administrator user role can view the security logs. If extended logging is disabled, you can only view system events, regardless of RBAC user role. To view security logs, use the `show logging` command.

Example of the `show logging auditlog` Command

For information about the logging extended command, see Enabling Audit and Security Logs

```
Dell#show logging auditlog
May 12 12:20:25: Dell#: %CLI-6-logging extended by admin from vty0 (10.14.1.98)
May 12 12:20:42: Dell#: %CLI-6-configure terminal by admin from vty0
(10.14.1.98)
May 12 12:20:42: Dell#: %CLI-6-service timestamps log datetime by admin from
vty0 (10.14.1.98)
```

Example of the `show logging` Command for Security

For information about the logging extended command, see Enabling Audit and Security Logs

```
Dell#show logging
Jun 10 04:23:40: %STKUNIT0-M:CP %SEC-5-LOGIN_SUCCESS: Login successful for
user admin on line vty0 ( 10.14.1.91 )
```

Clearing Audit Logs

To clear audit logs, use the `clear logging auditlog` command in Exec mode. When RBAC is enabled, only the system administrator user role can issue this command.

Example of the clear logging auditlog Command

```
Dell# clear logging auditlog
```

Configuring Logging Format

To display syslog messages in a RFC 3164 or RFC 5424 format, use the `logging version [0 | 1}` command in CONFIGURATION mode. By default, the system log version is set to `0`.

The following describes the two log messages formats:

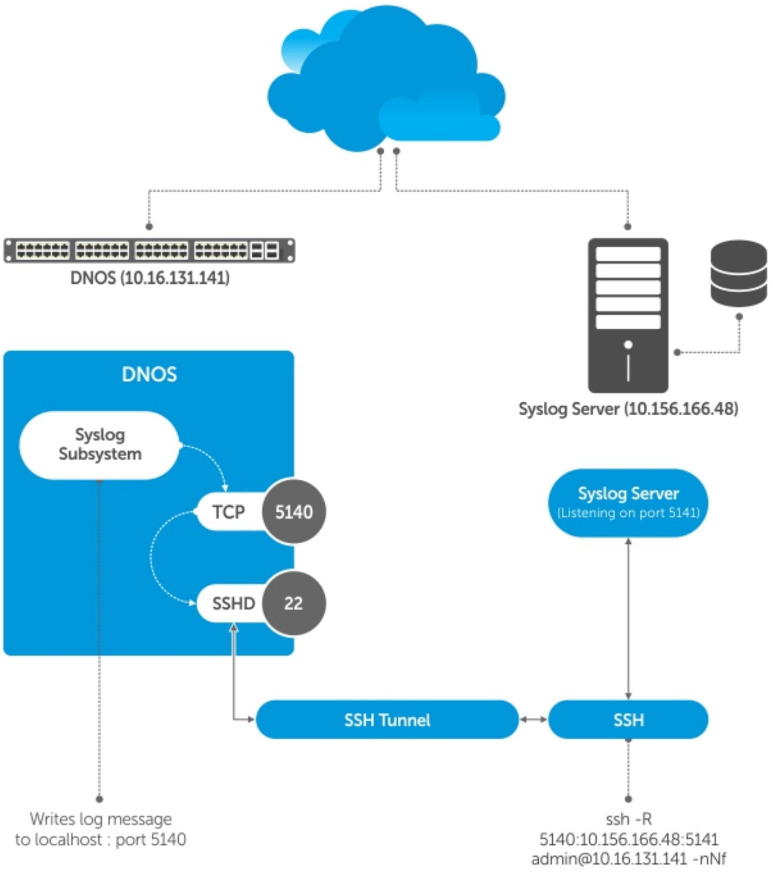
- **0** – Displays syslog messages format as described in RFC 3164, The BSD syslog Protocol
- **1** – Displays syslog message format as described in RFC 5424, The SYSLOG Protocol

Example of Configuring the Logging Message Format

```
Dell(conf)#logging version ?
<0-1> Select syslog version (default = 0)
Dell(conf)#logging version 1
```

Setting Up a Secure Connection to a Syslog Server

You can use reverse tunneling with the port forwarding to securely connect to a syslog server.


DNOS (10.16.131.141)
DNOS
Syslog Subsystem
TCP
5140
SSHD
22
Syslog Server (10.156.166.48)
Syslog Server
(Listening on port 5141)
SSH Tunnel
SSH
Writes log message to localhost : port 5140

```
ssh -R
5140:10.156.166.48:5141
admin@10.16.131.141 -nNf
```

Pre-requisites

To configure a secure connection from the switch to the syslog server:

1. On the switch, enable the SSH server

   ```
   Dell(conf)#ip ssh server enable
   ```

2. On the syslog server, create a reverse SSH tunnel from the syslog server to FTOS switch, using following syntax:

   ```
   ssh -R <remote port>:<syslog server>:<syslog server listen port>
   user@remote_host -nNf
   ```

   In the following example the syslog server IP address is `10.156.166.48` and the listening port is `5141`. The switch IP address is `10.16.131.141` and the listening port is `5140`

   ```
   ssh -R 5140:10.156.166.48:5141 admin@10.16.131.141 -nNf
   ```
3. Configure logging to a local host. *locahost* is "127.0.0.1" or "::1".

   If you do not, the system displays an error when you attempt to enable role-based only AAA authorization.

   ```
   Dell(conf)# logging localhost tcp port
   Dell(conf)#logging 127.0.0.1 tcp 5140
   ```

Log Messages in the Internal Buffer

All error messages, except those beginning with `%BOOTUP (Message)`, are log in the internal buffer.

For example, `%BOOTUP:RPM0:CP %PORTPIPE-INIT-SUCCESS: Portpipe 0 enabled`

Configuration Task List for System Log Management

There are two configuration tasks for system log management:

- Disable System Logging
- Send System Messages to a Syslog Server

Disabling System Logging

By default, logging is enabled and log messages are sent to the logging buffer, all terminal lines, the console, and the syslog servers.
To disable system logging, use the following commands.

- Disable all logging except on the console.

  CONFIGURATION mode

  ```
  no logging on
  ```
- Disable logging to the logging buffer.

  CONFIGURATION mode

  ```
  no logging buffer
  ```
- Disable logging to terminal lines.

  CONFIGURATION mode

  ```
  no logging monitor
  ```
- Disable console logging.

  CONFIGURATION mode

  ```
  no logging console
  ```

Sending System Messages to a Syslog Server

To send system messages to a specified syslog server, use the following command. The following syslog standards are supported: RFC 5424 The SYSLOG Protocol, R.Gerhards and Adiscon GmbH, March 2009, obsoletes RFC 3164 and RFC 5426 Transmission of Syslog Messages over UDP.

- Specify the server to which you want to send system messages. You can configure up to eight syslog servers.

  CONFIGURATION mode

  ```
  logging {ip-address | ipv6-address | hostname} {{udp {port}} | {tcp {port}}}
  ```

  In the release 9.4.(0.0), exporting of Syslogs to external servers that are connected through different VRFs is supported.

Configuring a UNIX System as a Syslog Server

To configure a UNIX System as a syslog server, use the following command.

- Configure a UNIX system as a syslog server by adding the following lines to */etc/syslog.conf* on the UNIX system and assigning write permissions to the file.
  - Add line on a 4.1 BSD UNIX system. `local7.debugging /var/log/ftos.log`
  - Add line on a 5.7 SunOS UNIX system. `local7.debugging /var/adm/ftos.log`

In the previous lines, `local7` is the logging facility level and debugging is the severity level.

Changing System Logging Settings

You can change the default settings of the system logging by changing the severity level and the storage location.

The default is to log all messages up to debug level, that is, all system messages. By changing the severity level in the logging commands, you control the number of system messages logged.

To specify the system logging settings, use the following commands.

- Specify the minimum severity level for logging to the logging buffer.

  CONFIGURATION mode

  ```
  logging buffered level
  ```
- Specify the minimum severity level for logging to the console.

  CONFIGURATION mode

  ```
  logging console level
  ```
- Specify the minimum severity level for logging to terminal lines.

  CONFIGURATION mode

  ```
  logging monitor level
  ```
- Specify the minimum severity level for logging to a syslog server.

  CONFIGURATION mode

  ```
  logging trap level
  ```

- Specify the minimum severity level for logging to the syslog history table.

  CONFIGURATION mode

  ```
  logging history level
  ```
- Specify the size of the logging buffer.

  CONFIGURATION mode

  ```
  logging buffered size
  ```

  NOTE: When you decrease the buffer size, Dell Networking OS deletes all messages stored in the buffer. Increasing the buffer size does not affect messages in the buffer.
- Specify the number of messages that Dell Networking OS saves to its logging history table.

  CONFIGURATION mode

  ```
  logging history size size
  ```

To view the logging buffer and configuration, use the `show logging` command in EXEC privilege mode, as shown in the example for Display the Logging Buffer and the Logging Configuration.

To view the logging configuration, use the `show running-config logging` command in privilege mode, as shown in the example for Configure a UNIX Logging Facility Level.

Display the Logging Buffer and the Logging Configuration

To display the current contents of the logging buffer and the logging settings for the system, use the `show logging` command in EXEC privilege mode. When RBAC is enabled, the security logs are filtered based on the user roles. Only the security administrator and system administrator can view the security logs.

Example of the `show logging` Command

```
Dell#show logging
syslog logging: enabled
   Console logging: level Debugging
   Monitor logging: level Debugging
   Buffer logging: level Debugging, 40 Messages Logged, Size (40960 bytes)
   Trap logging: level Informational
%IRC-6-IRC_COMMUP: Link to peer RPM is up
%RAM-6-RAM_TASK: RPM1 is transitioning to Primary RPM.
%RPM-2-MSG:CP1 %POLLMGR-2-MMC_STATE: External flash disk missing in 'slot0:'
%CHMGR-5-CARDDETECTED: Line card 0 present
%CHMGR-5-CARDDETECTED: Line card 2 present
%CHMGR-5-CARDDETECTED: Line card 4 present
%CHMGR-5-CARDDETECTED: Line card 5 present
%CHMGR-5-CARDDETECTED: Line card 8 present
%CHMGR-5-CARDDETECTED: Line card 10 present
%CHMGR-5-CARDDETECTED: Line card 12 present
%TSM-6-SFM_DISCOVERY: Found SFM 0
%TSM-6-SFM_DISCOVERY: Found SFM 1
%TSM-6-SFM_DISCOVERY: Found SFM 2
%TSM-6-SFM_DISCOVERY: Found SFM 3
%TSM-6-SFM_DISCOVERY: Found SFM 4
%TSM-6-SFM_DISCOVERY: Found SFM 5
%TSM-6-SFM_DISCOVERY: Found SFM 6
%TSM-6-SFM_DISCOVERY: Found SFM 7
%TSM-6-SFM_SWITCHFAB_STATE: Switch Fabric: UP
%TSM-6-SFM_DISCOVERY: Found SFM 8
%TSM-6-SFM_DISCOVERY: Found 9 SFMs
```

```
%CHMGR-5-CHECKIN: Checkin from line card 5 (type EX1YB, 1 ports)
%TSM-6-PORT_CONFIG: Port link status for LC 5 => portpipe 0: OK portpipe 1: N/A
%CHMGR-5-LINECARDUP: Line card 5 is up
%CHMGR-5-CHECKIN: Checkin from line card 12 (type S12YC12, 12 ports)
%TSM-6-PORT_CONFIG: Port link status for LC 12 => portpipe 0: OK portpipe 1: N/A
%CHMGR-5-LINECARDUP: Line card 12 is up
%IFMGR-5-CSTATE_UP: changed interface Physical state to up: So 12/8
%IFMGR-5-CSTATE_DN: changed interface Physical state to down: So 12/8
```

To view any changes made, use the `show running-config logging` command in EXEC privilege mode, as shown in the example for Configure a UNIX Logging Facility Level.

Configuring a UNIX Logging Facility Level

You can save system log messages with a UNIX system logging facility.

To configure a UNIX logging facility level, use the following command.

- Specify one of the following parameters.

  CONFIGURATION mode

  `logging facility [`*`facility-type`*`]`

  - `auth` (for authorization messages)
  - `cron` (for system scheduler messages)
  - `daemon` (for system daemons)
  - `kern` (for kernel messages)
  - `local0` (for local use)
  - `local1` (for local use)
  - `local2` (for local use)
  - `local3` (for local use)
  - `local4` (for local use)
  - `local5` (for local use)
  - `local6` (for local use)
  - `local7` (for local use)
  - `lpr` (for line printer system messages)
  - `mail` (for mail system messages)
  - `news` (for USENET news messages)
  - `sys9` (system use)
  - `sys10` (system use)
  - `sys11` (system use)
  - `sys12` (system use)
  - `sys13` (system use)
  - `sys14` (system use)
  - `syslog` (for syslog messages)
  - `user` (for user programs)
  - `uucp` (UNIX to UNIX copy protocol)

Example of the `show running-config logging` Command

To view nondefault settings, use the `show running-config logging` command in EXEC mode.

```
Dell#show running-config logging
!
logging buffered 524288 debugging
service timestamps log datetime msec
service timestamps debug datetime msec
!
logging trap debugging
logging facility user
logging source-interface Loopback 0
logging 10.10.10.4
Dell#
```

Synchronizing Log Messages

You can configure Dell Networking OS to filter and consolidate the system messages for a specific line by synchronizing the message output.

Only the messages with a severity at or below the set level appear. This feature works on the terminal and console connections available on the system.

1. Enter LINE mode.

   CONFIGURATION mode

   ```
   line {console 0 | vty number [end-number] | aux 0}
   ```

   Configure the following parameters for the virtual terminal lines:

   - *`number`*: the range is from zero (0) to 8.
   - *`end-number`*: the range is from 1 to 8.

   You can configure multiple virtual terminals at one time by entering a *number* and an *end-number*.

2. Configure a level and set the maximum number of messages to print.

   LINE mode

   ```
   logging synchronous [level severity-level | all] [limit]
   ```

   Configure the following optional parameters:

   - `level` *`severity-level`*: the range is from 0 to 7. The default is **2**. Use the `all` keyword to include all messages.
   - `limit`: the range is from 20 to 300. The default is **20**.

To view the logging synchronous configuration, use the `show config` command in LINE mode.

Enabling Timestamp on Syslog Messages

By default, syslog messages do not include a time/date stamp stating when the error or message was created.

To enable timestamp, use the following command.

- Add timestamp to syslog messages.

CONFIGURATION mode

```
service timestamps [log | debug] [datetime [localtime] [msec] [show-timezone]
| uptime]
```

Specify the following optional parameters:

- You can add the keyword `localtime` to include the `localtime`, `msec`, and `show-timezone`. If you do not add the keyword `localtime`, the time is UTC.
- `uptime`: To view time since last boot.

If you do not specify a parameter, Dell Networking OS configures `uptime`.

To view the configuration, use the `show running-config logging` command in EXEC privilege mode.

To disable time stamping on syslog messages, use the `no service timestamps [log | debug]` command.

File Transfer Services

With Dell Networking OS, you can configure the system to transfer files over the network using the file transfer protocol (FTP).

One FTP application is copying the system image files over an interface on to the system; however, FTP is not supported on virtual local area network (VLAN) interfaces.

In the release 9.4.(0.0), FTP and TFTP services are enhanced to support the VRF-aware functionality. If you want the FTP or TFTP server to use a VRF table that is attached to an interface, you must configure the FTP or TFTP server to use a specific routing table. You can use the `ip ftp vrf` *`vrf-name`* or `ip tftp vrf` *`vrf-name`* command to inform the FTP or TFTP server to use a specific routing table. After you configure this setting, the VRF table is used to look up the destination address. However, these changes are backward-compatible and do not affect existing behavior; meaning, you can still use the `source-interface` command to communicate with a particular interface even if no VRF is configured on that interface.

For more information about FTP, refer to RFC 959, *File Transfer Protocol*.

NOTE: To transmit large files, Dell Networking recommends configuring the switch as an FTP server.

Configuration Task List for File Transfer Services

The configuration tasks for file transfer services are:

- Enable FTP Server (mandatory)
- Configure FTP Server Parameters (optional)
- Configure FTP Client Parameters (optional)

Enabling the FTP Server

To enable the system as an FTP server, use the following command.

To view FTP configuration, use the `show running-config ftp` command in EXEC privilege mode.

- Enable FTP on the system.

CONFIGURATION mode

```
ftp-server enable
```

Example of Viewing FTP Configuration

```
Dell#show running ftp
!
ftp-server enable
ftp-server username nairobi password 0 zanzibar
Dell#
```

Configuring FTP Server Parameters

After you enable the FTP server on the system, you can configure different parameters.
To specify the system logging settings, use the following commands.

- Specify the directory for users using FTP to reach the system.
  CONFIGURATION mode

  ```
  ftp-server topdir dir
  ```

  The default is the internal flash directory.
- Specify a user name for all FTP users and configure either a plain text or encrypted password.
  CONFIGURATION mode

  ```
  ftp-server username username password [encryption-type] password
  ```

  Configure the following optional and required parameters:
  - *`username`*: enter a text string.
  - *`encryption-type`*: enter `0` for plain text or `7` for encrypted text.
  - *`password`*: enter a text string.

NOTE: You cannot use the `change directory (cd)` command until you have configured `ftp-server topdir`.

To view the FTP configuration, use the `show running-config ftp` command in EXEC privilege mode.

Configuring FTP Client Parameters

To configure FTP client parameters, use the following commands.

- Enter the following keywords and slot/port or number information:
  - For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
  - For a loopback interface, enter the keyword `loopback` then a number between 0 and 16383.
  - For a port channel interface, enter the keywords `port-channel` then a number from 1 to 255 for TeraScale and ExaScale.
  - For a SONET interface, enter the keyword `sonet` then the slot/port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.

  - For a VLAN interface, enter the keyword `vlan` then a number from 1 to 4094.
  - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.

  CONFIGURATION mode

  `ip ftp source-interface` *`interface`*
- Configure a password.

  CONFIGURATION mode

  `ip ftp password` *`password`*
- Enter a username to use on the FTP client.

  CONFIGURATION mode

  `ip ftp username` *`name`*

To view the FTP configuration, use the `show running-config ftp` command in EXEC privilege mode, as shown in the example for Enable FTP Server.

Terminal Lines

You can access the system remotely and restrict access to the system by creating user profiles.

Terminal lines on the system provide different means of accessing the system. The console line (console) connects you through the console port in the route processor modules (RPMs). The virtual terminal lines (VTYs) connect you through Telnet to the system. The auxiliary line (aux) connects secondary devices such as modems.

Denying and Permitting Access to a Terminal Line

Dell Networking recommends applying only standard access control lists (ACLs) to deny and permit access to VTY lines.

- Layer 3 ACLs deny all traffic that is not explicitly permitted, but in the case of VTY lines, an ACL with no rules does not deny traffic.
- You cannot use the `show ip accounting access-list` command to display the contents of an ACL that is applied only to a VTY line.

To apply an IP ACL to a line, Use the following command.

- Apply an ACL to a VTY line.

  LINE mode

  `ip access-class` *`access-list`*

Example of an ACL that Permits Terminal Access

To view the configuration, use the `show config` command in LINE mode.

```
Dell(config-std-nacl)#show config
!
ip access-list standard myvtyacl
  seq 5 permit host 10.11.0.1
Dell(config-std-nacl)#line vty 0
Dell(config-line-vty)#show config
```

```
line vty 0
  access-class myvtyacl
```

Dell Networking OS Behavior: Prior to Dell Networking OS version 7.4.2.0, in order to deny access on a VTY line, apply an ACL and accounting, authentication, and authorization (AAA) to the line. Then users are denied access only after they enter a username and password. Beginning in Dell Networking OS version 7.4.2.0, only an ACL is required, and users are denied access before they are prompted for a username and password.

Configuring Login Authentication for Terminal Lines

You can use any combination of up to six authentication methods to authenticate a user on a terminal line.
A combination of authentication methods is called a method list. If the user fails the first authentication method, Dell Networking OS prompts the next method until all methods are exhausted, at which point the connection is terminated. The available authentication methods are:

`enable`	Prompt for the enable password.
`line`	Prompt for the password you assigned to the terminal line. Configure a password for the terminal line to which you assign a method list that contains the line authentication method. Configure a password using the `password` command from LINE mode.
`local`	Prompt for the system username and password.
`none`	Do not authenticate the user.
`radius`	Prompt for a username and password and use a RADIUS server to authenticate.
`tacacs+`	Prompt for a username and password and use a TACACS+ server to authenticate.

1. Configure an authentication method list. You may use a mnemonic name or use the keyword default. The default authentication method for terminal lines is **local** and the default method list is **empty**.

   CONFIGURATION mode

   ```
   aaa authentication login {method-list-name | default} [method-1] [method-2]
   [method-3] [method-4] [method-5] [method-6]
   ```

2. Apply the method list from Step 1 to a terminal line.

   CONFIGURATION mode

   ```
   login authentication {method-list-name | default}
   ```

3. If you used the line authentication method in the method list you applied to the terminal line, configure a password for the terminal line.

   LINE mode

   ```
   password
   ```

Example of Terminal Line Authentication

In the following example, VTY lines 0-2 use a single authentication method, line.

```
Dell(conf)#aaa authentication login myvtymethodlist line
Dell(conf)#line vty 0 2
Dell(config-line-vty)#login authentication myvtymethodlist
Dell(config-line-vty)#password myvtypassword
```

```
Dell(config-line-vty)#show config
line vty 0
  password myvtypassword
login authentication myvtymethodlist
line vty 1
  password myvtypassword
login authentication myvtymethodlist
line vty 2
  password myvtypassword
login authentication myvtymethodlist
Dell(config-line-vty)#
```

Setting Time Out of EXEC Privilege Mode

EXEC time-out is a basic security feature that returns Dell Networking OS to EXEC mode after a period of inactivity on the terminal lines.
To set time out, use the following commands.

- Set the number of minutes and seconds. The default is **10 minutes** on the console and `30 minutes` on VTY. Disable EXEC time out by setting the time-out period to `0`.

  LINE mode

  ```
  exec-timeout minutes [seconds]
  ```
- Return to the default time-out values.

  LINE mode

  ```
  no exec-timeout
  ```

Example of Setting the Time Out Period for EXEC Privilege Mode

The following example shows how to set the time-out period and how to view the configuration using the `show config` command from LINE mode.

```
Dell(conf)#line con 0
Dell(config-line-console)#exec-timeout 0
Dell(config-line-console)#show config
line console 0
  exec-timeout 0 0
Dell(config-line-console)#
```

Using Telnet to get to Another Network Device

To telnet to another device, use the following commands.

NOTE: On the platform, the system allows 120 Telnet sessions per minute, allowing the login and logout of 10 Telnet sessions, 12 times in a minute. If the system reaches this non-practical limit, the Telnet service is stopped for 10 minutes. You can use console and SSH service to access the system during downtime.

- Telnet to the peer RPM. You do not need to configure the management port on the peer RPM to be able to telnet to it.

  EXEC Privilege mode

  ```
  telnet-peer-rpm
  ```
- Telnet to a device with an IPv4 or IPv6 address.

EXEC Privilege

`telnet [ip-address]`

If you do not enter an IP address, Dell Networking OS enters a Telnet dialog that prompts you for one.

Enter an IPv4 address in dotted decimal format (A.B.C.D).

Enter an IPv6 address in the format 0000:0000:0000:0000:0000:0000:0000:0000. Elision of zeros is supported.

Example of the `telnet` Command for Device Access

```
Dell# telnet 10.11.80.203
Trying 10.11.80.203...
Connected to 10.11.80.203.
Exit character is '^]'.
Login:
Login: admin
Password:
Dell>exit
Dell#telnet 2200:2200:2200:2200:2200::2201
Trying 2200:2200:2200:2200:2200::2201...
Connected to 2200:2200:2200:2200:2200::2201.
Exit character is '^]'.
FreeBSD/i386 (freebsd2.force10networks.com) (ttyp1)
login: admin
Dell#
```

Lock CONFIGURATION Mode

Dell Networking OS allows multiple users to make configurations at the same time. You can lock CONFIGURATION mode so that only one user can be in CONFIGURATION mode at any time (Message 2).

You can set two types of lockst: auto and manual.

- Set auto-lock using the `configuration mode exclusive auto` command from CONFIGURATION mode. When you set auto-lock, every time a user is in CONFIGURATION mode, all other users are denied access. This means that you can exit to EXEC Privilege mode, and re-enter CONFIGURATION mode without having to set the lock again.
- Set manual lock using the `configure terminal lock` command from CONFIGURATION mode. When you configure a manual lock, which is the default, you must enter this command each time you want to enter CONFIGURATION mode and deny access to others.

Viewing the Configuration Lock Status

If you attempt to enter CONFIGURATION mode when another user has locked it, you may view which user has control of CONFIGURATION mode using the `show configuration lock` command from EXEC Privilege mode.

You can then send any user a message using the `send` command from EXEC Privilege mode. Alternatively, you can clear any line using the `clear` command from EXEC Privilege mode. If you clear a console session, the user is returned to EXEC mode.

Example of Locking CONFIGURATION Mode for Single-User Access

```
Dell(conf)#configuration mode exclusive auto
BATMAN(conf)#exit
3d23h35m: %RPM0-P:CP %SYS-5-CONFIG_I: Configured from console by console

Dell#config
! Locks configuration mode exclusively.
Dell(conf)#
```

If another user attempts to enter CONFIGURATION mode while a lock is in place, the following appears on their terminal (message 1): `% Error: User "" on line console0 is in exclusive configuration mode.`

If *any* user is already in CONFIGURATION mode when while a lock is in place, the following appears on their terminal (message 2): `% Error: Can't lock configuration mode exclusively since the following users are currently configuring the system: User "admin" on line vty1 ( 10.1.1.1 ).`

NOTE: The CONFIGURATION mode lock corresponds to a VTY session, not a user. Therefore, if you configure a lock and then exit CONFIGURATION mode, and another user enters CONFIGURATION mode, when you attempt to re-enter CONFIGURATION mode, you are denied access even though you are the one that configured the lock.

NOTE: If your session times out and you return to EXEC mode, the CONFIGURATION mode lock is unconfigured.

Restoring the Factory Default Settings

Restoring the factory-default settings deletes the existing NVRAM settings, startup configuration, and all configured settings such as, stacking or fanout.

S6000MXL Switch

To restore the factory default settings, use the `restore factory-defaults stack-unit {0-5 | all} {clear-all | nvram}` command in EXEC Privilege mode.

CAUTION: There is no undo for this command.

Important Points to Remember

- When you restore all the units in a stack, these units are placed in standalone mode.
- When you restore a single unit in a stack, only that unit is placed in standalone mode. No other units in the stack are affected.
- When you restore the units in standalone mode, the units remain in standalone mode after the restoration.
- After the restore is complete, the units power cycle immediately.

The following example illustrates the **restore factory-defaults command** to restore the factory default settings.

```
Dell#restore factory-defaults stack-unit 0 nvram

    ***********************************************************************
    *  Warning - Restoring factory defaults will delete the existing      *
    *  persistent settings (stacking, fanout, etc.)                    *
    *  After restoration the unit(s) will be powercycled immediately.    *
    *  Proceed with caution !                                          *
    ***********************************************************************

Proceed with factory settings? Confirm [yes/no]:yes

-- Restore status --
Unit   Nvram     Config
------------------------
  0    Success

Power-cycling the unit(s).
....
```

Restoring Factory Default Environment Variables

The Boot line determines the location of the image that is used to boot up the chassis after restoring factory default settings. Ideally, these locations contain valid images, using which the chassis boots up.

While restoring factory-default settings, you can either use a flash boot procedure or a network boot procedure to boot the device.

When you use the flash boot procedure to boot the device, the boot loader checks if the primary or the secondary partition contains a valid image. If the primary partition contains a valid image, then the primary boot line is set to A: and the secondary and default boot lines are set to a Null String. If the secondary partition contains a valid image, then the primary boot line is set to B: and the secondary and default boot lines are set to a Null String. If both the partitions contain invalid images, then primary, secondary, and default boot line values are set to a Null string.

When you use the Network boot procedure to boot the device, the boot loader checks if the primary partition contains a valid image. If a valid image exists on the primary partition and the secondary partition does not contain a valid image, then the primary boot line is set to A: and the secondary and default boot lines are set to a Null string. If the secondary partition also contains a valid image, then the primary boot line value is set to the partition that is configured to be used to boot the device in a network failure scenario. The secondary and default boot line values are set to a Null string.

Important Points to Remember

- The Chassis remains in boot prompt if none of the partitions contain valid images.
- To enable TFTP boot after restoring factory default settings, you must stop the boot process in BLI.

In case the system fails to reload the image from the partition, perform the following steps:

1. Power-cycle the chassis (pull the power cord and reinsert it).
2. Press esc key to abort the boot process (while the system prompts to)

   You enter BLI immediately, as indicated by the BOOT_USER # prompt.

   press any key
3. Assign the new location of the FTOS image to be used when the system reloads.

   To boot from flash partition A:

```
BOOT_USER # boot change primary

boot device : flash

file name : systema

BOOT_USER #
```

To boot from flash partition B:

```
BOOT_USER # boot change primary

boot device : flash

file name : systemb

BOOT_USER #
```

To boot from network:

```
BOOT_USER # boot change primary

boot device : tftp

file name : FTOS-SI-9-5-0-169.bin

Server IP address : 10.16.127.35

BOOT_USER #
```

4. Assign an IP address and netmask to the Management Ethernet interface.

   ```
   BOOT_USER # interface management ethernet ip address ip_address_with_mask
   ```

   For example, *10.16.150.106/16*.
5. Assign an IP address as the default gateway for the system.

   ```
   default-gateway gateway_ip_address
   ```

   For example, *10.16.150.254*.
6. The environment variables are auto saved.
7. Reload the system.

   ```
   BOOT_USER # reload
   ```

5

802.1X

802.1X is supported on the S6000 platform.

802.1X is a method of port security. A device connected to a port that is enabled with 802.1X is disallowed from sending or receiving packets on the network until its identity can be verified (through a username and password, for example). This feature is named for its IEEE specification.

802.1X employs extensible authentication protocol (EAP) to transfer a device’s credentials to an authentication server (typically RADIUS) using a mandatory intermediary network access device, in this case, a Dell Networking switch. The network access device mediates all communication between the end-user device and the authentication server so that the network remains secure. The network access device uses EAP-over-Ethernet (EAPOL) to communicate with the end-user device and EAP-over-RADIUS to communicate with the server.

NOTE: The Dell Networking Operating System (OS) supports 802.1X with EAP-MD5, EAP-OTP, EAP-TLS, EAP-TTLS, PEAPv0, PEAPv1, and MS-CHAPv2 with PEAP.

The following figures show how the EAP frames are encapsulated in Ethernet and RADIUS frames.

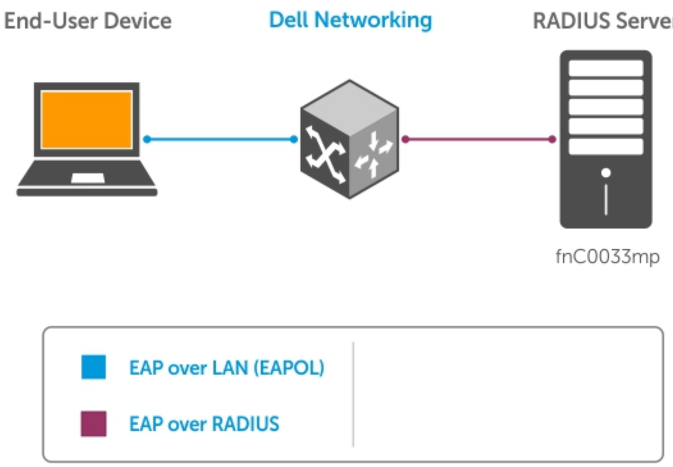


Figure 2. EAP Frames Encapsulated in Ethernet and RADUIS

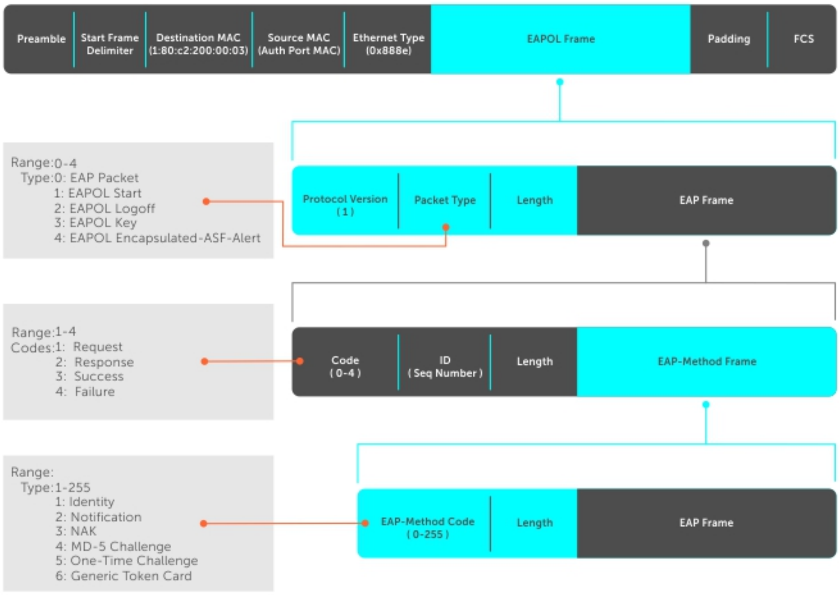

Preamble
Start Frame Delimiter
Destination MAC (1:80:c2:200:00:03)
Source MAC (Auth Port MAC)
Ethernet Type (0x888e)
EAPOL Frame
Padding
FCS
Range:0-4
Type:0: EAP Packet
1: EAPOL Start
2: EAPOL Logoff
3: EAPOL Key
4: EAPOL Encapsulated-ASF-Alert
Protocol Version (1)
Packet Type
Length
EAP Frame
Range:1-4
Codes:1: Request
2: Response
3: Success
4: Failure
Code (0-4)
ID (Seq Number)
Length
EAP-Method Frame
Range:
Type:1-255
1: Identity
2: Notification
3: NAK
4: MD-5 Challenge
5: One-Time Challenge
6: Generic Token Card
EAP-Method Code (0-255)
Length
EAP Frame

Figure 3. EAP Frames Encapsulated in Ethernet and RADUIS

The authentication process involves three devices:

- The device attempting to access the network is the **supplicant**. The supplicant is not allowed to communicate on the network until the authenticator authorizes the port. It can only communicate with the authenticator in response to 802.1X requests.
- The device with which the supplicant communicates is the **authenticator**. The authenticator is the gate keeper of the network. It translates and forwards requests and responses between the authentication server and the supplicant. The authenticator also changes the status of the port based on the results of the authentication process. The Dell Networking switch is the authenticator.
- The authentication-server selects the authentication method, verifies the information the supplicant provides, and grants it network access privileges.

Ports can be in one of two states:

- Ports are in an **unauthorized** state by default. In this state, non-802.1X traffic cannot be forwarded in or out of the port.
- The authenticator changes the port state to authorized if the server can authenticate the supplicant. In this state, network traffic can be forwarded normally.

NOTE: The Dell Networking switches place 802.1X-enabled ports in the unauthorized state by default.

The Port-Authentication Process

The authentication process begins when the authenticator senses that a link status has changed from down to up:

1. When the authenticator senses a link state change, it requests that the supplicant identify itself using an EAP Identity Request frame.
2. The supplicant responds with its identity in an EAP Response Identity frame.

3. The authenticator decapsulates the EAP response from the EAPOL frame, encapsulates it in a RADIUS Access-Request frame and forwards the frame to the authentication server.
4. The authentication server replies with an Access-Challenge frame. The Access-Challenge frame requests that the supplicant prove that it is who it claims to be, using a specified method (an EAP-Method). The challenge is translated and forwarded to the supplicant by the authenticator.
5. The supplicant can negotiate the authentication method, but if it is acceptable, the supplicant provides the Requested Challenge information in an EAP response, which is translated and forwarded to the authentication server as another Access-Request frame.
6. If the identity information provided by the supplicant is valid, the authentication server sends an Access-Accept frame in which network privileges are specified. The authenticator changes the port state to authorized and forwards an EAP Success frame. If the identity information is invalid, the server sends an Access-Reject frame. If the port state remains unauthorized, the authenticator forwards an EAP Failure frame.

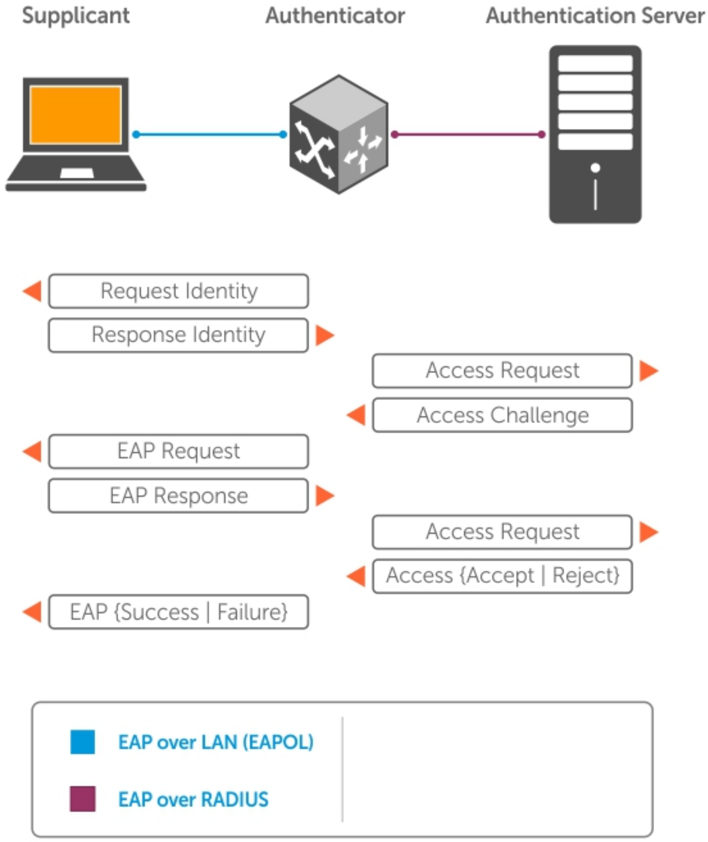

Supplicant
Authenticator
Authentication Server
Request Identity
Response Identity
Access Request
Access Challenge
EAP Request
EAP Response
Access Request
Access {Accept | Reject}
EAP {Success | Failure}
EAP over LAN (EAPOL)
EAP over RADIUS

Figure 4. EAP Port-Authentication

EAP over RADIUS

802.1X uses RADIUS to shuttle EAP packets between the authenticator and the authentication server, as defined in RFC 3579.

EAP messages are encapsulated in RADIUS packets as a type of attribute in Type, Length, Value (TLV) format. The Type value for EAP messages is 79.

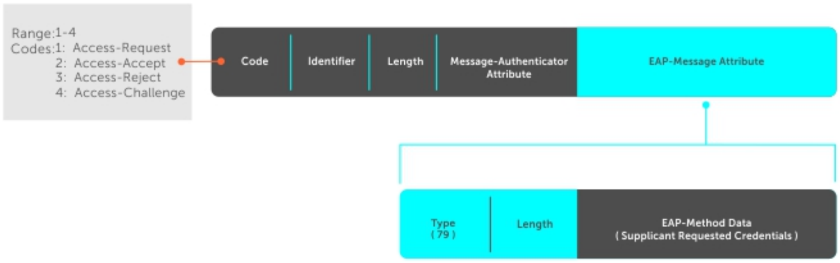

Range:1-4
Codes:1: Access-Request
2: Access-Accept
3: Access-Reject
4: Access-Challenge
Code
Identifier
Length
Message-Authenticator Attribute
EAP-Message Attribute
Type (79)
Length
EAP-Method Data (Supplicant Requested Credentials)

Figure 5. EAP Over RADIUS

RADIUS Attributes for 802.1 Support

Dell Networking systems include the following RADIUS attributes in all 802.1X-triggered Access-Request messages:

Attribute 31	**Calling-station-id**: relays the supplicant MAC address to the authentication server.
Attribute 41	**NAS-Port-Type**: NAS-port physical port type. 15 indicates Ethernet.
Attribute 61	**NAS-Port**: the physical port number by which the authenticator is connected to the supplicant.
Attribute 81	**Tunnel-Private-Group-ID**: associate a tunneled session with a particular group of users.

Configuring 802.1X

Configuring 802.1X on a port is a one-step process.

For more information, refer to Enabling 802.1X.

Related Configuration Tasks

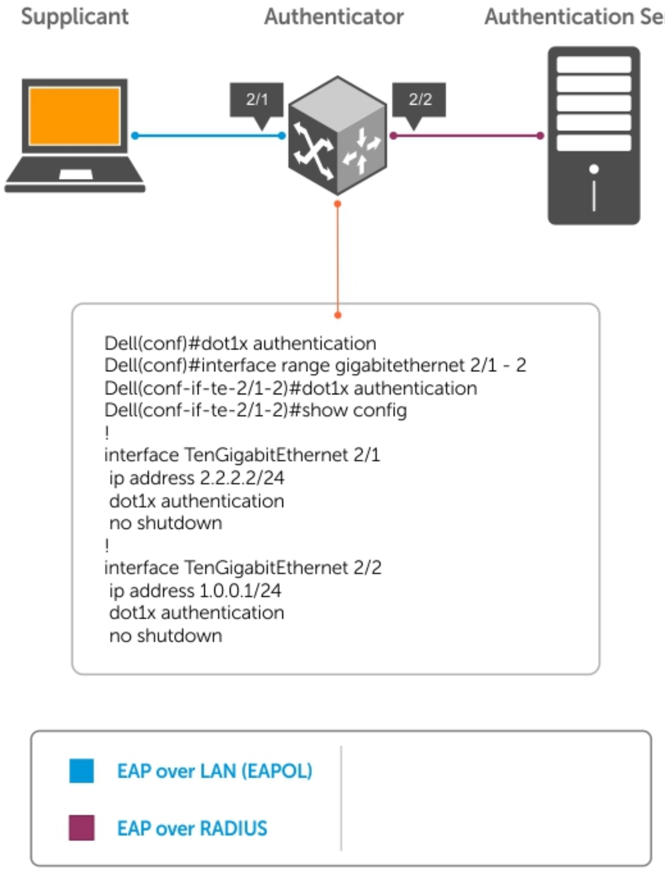
- Configuring Request Identity Re-Transmissions
- Forcibly Authorizing or Unauthorizing a Port
- Re-Authenticating a Port
- Configuring Timeouts
- Configuring a Guest VLAN
- Configuring an Authentication-Fail VLAN

Important Points to Remember

- Dell Networking OS supports 802.1X with EAP-MD5, EAP-OTP, EAP-TLS, EAP-TTLS, PEAPv0, PEAPv1, and MS-CHAPv2 with PEAP.
- All platforms support only RADIUS as the authentication server.
- If the primary RADIUS server becomes unresponsive, the authenticator begins using a secondary RADIUS server, if configured.
- 802.1X is not supported on port-channels or port-channel members.

Enabling 802.1X

Enable 802.1X globally.



```
Dell(conf)#dot1x authentication
Dell(conf)#interface range gigabitethernet 2/1 - 2
Dell(conf-if-te-2/1-2)#dot1x authentication
Dell(conf-if-te-2/1-2)#show config
!
interface TenGigabitEthernet 2/1
 ip address 2.2.2.2/24
 dot1x authentication
 no shutdown
!
interface TenGigabitEthernet 2/2
 ip address 1.0.0.1/24
 dot1x authentication
 no shutdown
```

Figure 6. 802.1X Enabled

1. Enable 802.1X globally.

   CONFIGURATION mode

```
dot1x authentication
```

2. Enter INTERFACE mode on an interface or a range of interfaces.

   INTERFACE mode

   `interface [`*`range`*`]`

3. Enable 802.1X on the supplicant interface only.

   INTERFACE mode

   ```
   dot1x authentication
   ```

Examples of Verifying that 802.1X is Enabled Globally and on an Interface

Verify that 802.1X is enabled globally and at the interface level using the `show running-config | find dot1x` command from EXEC Privilege mode.

In the following example, the bold lines show that 802.1X is enabled.

```
Dell#show running-config | find dot1x
dot1x authentication
!
[output omitted]
!
interface TenGigabitEthernet 2/1
 no ip address
 dot1x authentication
 no shutdown
!
Dell#
```

To view 802.1X configuration information for an interface, use the `show dot1x interface` command.

In the following example, the bold lines show that 802.1X is enabled on all ports unauthorized by default.

```
Dell#show dot1x interface TenGigabitEthernet 2/1

802.1x information on Te 2/1:
-----------------------------
Dot1x Status:           Enable
Port Control:           AUTO
Port Auth Status:       UNAUTHORIZED
Re-Authentication:      Disable
Untagged VLAN id:       None
Guest VLAN:             Disable
Guest VLAN id:          NONE
Auth-Fail VLAN:         Disable
Auth-Fail VLAN id:      NONE
Auth-Fail Max-Attempts: NONE
Mac-Auth-Bypass:        Disable
Mac-Auth-Bypass Only:   Disable
Tx Period:              30 seconds
Quiet Period:           60 seconds
ReAuth Max:             2
Supplicant Timeout:     30 seconds
Server Timeout:         30 seconds
Re-Auth Interval:       3600 seconds
Max-EAP-Req:            2
Host Mode:              SINGLE_HOST
Auth PAE State:         Initialize
Backend State:          Initialize
```

Configuring Request Identity Re-Transmissions

If the authenticator sends a Request Identity frame, but the supplicant does not respond, the authenticator waits 30 seconds and then re-transmits the frame.
The amount of time that the authenticator waits before re-transmitting and the maximum number of times that the authenticator re-transmits are configurable.

NOTE: There are several reasons why the supplicant might fail to respond; for example, the supplicant might have been booting when the request arrived or there might be a physical layer problem.

To configure re-transmissions, use the following commands.

- Configure the amount of time that the authenticator waits before re-transmitting an EAP Request Identity frame.

  INTERFACE mode

  ```
  dot1x tx-period number
  ```

  The range is from 1 to 65535 (1 year)

  The default is **30**.
- Configure a maximum number of times the authenticator re-transmits a Request Identity frame.

  INTERFACE mode

  ```
  dot1x max-eap-req number
  ```

  The range is from 1 to 10.

  The default is **2**.

The example in Configuring a Quiet Period after a Failed Authentication shows configuration information for a port for which the authenticator re-transmits an EAP Request Identity frame after 90 seconds and re-transmits a maximum of 10 times.

Configuring a Quiet Period after a Failed Authentication

If the supplicant fails the authentication process, the authenticator sends another Request Identity frame after 30 seconds by default, but you can configure this period.

NOTE: The quiet period (`dot1x quiet-period`) is a transmit interval for after a failed authentication; the Request Identity Re-transmit interval (`dot1x tx-period`) is for an unresponsive supplicant.

To configure a quiet period, use the following command.

- Configure the amount of time that the authenticator waits to re-transmit a Request Identity frame after a failed authentication.

  INTERFACE mode

  ```
  dot1x quiet-period seconds
  ```

  The range is from 1 to 65535.

  The default is **60 seconds**.

Example of Configuring and Verifying Port Authentication

The following example shows configuration information for a port for which the authenticator re-transmits an EAP Request Identity frame:

- after 90 seconds and a maximum of 10 times for an unresponsive supplicant
- re-transmits an EAP Request Identity frame

The bold lines show the new re-transmit interval, new quiet period, and new maximum re-transmissions.

```
FTOS(conf-if-range-Te-0/0)#dot1x tx-period 90
FTOS(conf-if-range-Te-0/0)#dot1x max-eap-req 10
FTOS(conf-if-range-Te-0/0)#dot1x quiet-period 120
FTOS#show dot1x interface TenGigabitEthernet 2/1
802.1x information on Te 2/1:
-----------------------------
Dot1x Status:          Enable
Port Control:          AUTO
Port Auth Status:      UNAUTHORIZED
Re-Authentication:   Disable
Untagged VLAN id:      None
Tx Period:             90 seconds
Quiet Period:       120 seconds
ReAuth Max:            2
Supplicant Timeout:    30 seconds
Server Timeout:        30 seconds
Re-Auth Interval:      3600 seconds
Max-EAP-Req:       10
Auth Type:             SINGLE_HOST
Auth PAE State:        Initialize
Backend State:         Initialize
```

Forcibly Authorizing or Unauthorizing a Port

IEEE 802.1X requires that a port can be manually placed into any of three states:

- **ForceAuthorized** — an authorized state. A device connected to this port in this state is never subjected to the authentication process, but is allowed to communicate on the network. Placing the port in this state is same as disabling 802.1X on the port.
- **ForceUnauthorized** — an unauthorized state. A device connected to a port in this state is never subjected to the authentication process and is not allowed to communicate on the network. Placing the port in this state is the same as shutting down the port. Any attempt by the supplicant to initiate authentication is ignored.
- **Auto** — an unauthorized state by default. A device connected to this port in this state is subjected to the authentication process. If the process is successful, the port is authorized and the connected device can communicate on the network. All ports are placed in the Auto state by default.

To set the port state, use the following command.

- Place a port in the ForceAuthorized, ForceUnauthorized, or Auto state.

  INTERFACE mode

  ```
  dot1x port-control {force-authorized | force-unauthorized | auto}
  ```

  The default state is **auto**.

Example of Placing a Port in Force-Authorized State and Viewing the Configuration

The example shows configuration information for a port that has been force-authorized.

The bold line shows the new port-control state.

```
Dell(conf-if-Te-0/0)#dot1x port-control force-authorized
Dell(conf-if-Te-0/0)#show dot1x interface TenGigabitEthernet 0/0

802.1x information on Te 0/0:
-----------------------------
Dot1x Status:          Enable
Port Control:       FORCE_AUTHORIZED
Port Auth Status:      UNAUTHORIZED
Re-Authentication:     Disable
Untagged VLAN id:      None
Tx Period:             90 seconds
Quiet Period:          120 seconds
ReAuth Max:            2
Supplicant Timeout:    30 seconds
Server Timeout:        30 seconds
Re-Auth Interval:      3600 seconds
Max-EAP-Req:           10
Auth Type:             SINGLE_HOST
Auth PAE State:        Initialize
Backend State:         Initialize
Auth PAE State:        Initialize
Backend State:         Initialize
```

Re-Authenticating a Port

You can configure the authenticator for periodic re-authentication.
After the supplicant has been authenticated, and the port has been authorized, you can configure the authenticator to re-authenticate the supplicant periodically. If you enable re-authentication, the supplicant is required to re-authenticate every 3600 seconds, but you can configure this interval. You can configure a maximum number of re-authentications as well.

To configure re-authentication time settings, use the following commands.

- Configure the authenticator to periodically re-authenticate the supplicant.
  INTERFACE mode

  `dot1x reauthentication [interval] *seconds*`

  The range is from 1 to 65535.

  The default is **3600**.
- Configure the maximum number of times that the supplicant can be re-authenticated.
  INTERFACE mode

  `dot1x reauth-max *number*`

  The range is from 1 to 10.

  The default is **2**.

Example of Re-Authenticating a Port and Verifying the Configuration

The bold lines show that re-authentication is enabled and the new maximum and re-authentication time period.

```
Dell(conf-if-Te-0/0)#dot1x reauthentication interval 7200
Dell(conf-if-Te-0/0)#dot1x reauth-max 10
Dell(conf-if-Te-0/0)#do show dot1x interface TenGigabitEthernet 0/0

802.1x information on Te 0/0:
-----------------------------
Dot1x Status:           Enable
Port Control:           FORCE_AUTHORIZED
Port Auth Status:       UNAUTHORIZED
Re-Authentication:      Enable
Untagged VLAN id:       None
Tx Period:              90 seconds
Quiet Period:           120 seconds
ReAuth Max:             10
Supplicant Timeout:     30 seconds
Server Timeout:         30 seconds
Re-Auth Interval:       7200 seconds
Max-EAP-Req:            10
Auth Type:              SINGLE_HOST
Auth PAE State:         Initialize
Backend State:          Initialize
Auth PAE State:         Initialize
Backend State:          Initialize
```

Configuring Timeouts

If the supplicant or the authentication server is unresponsive, the authenticator terminates the authentication process after 30 seconds by default. You can configure the amount of time the authenticator waits for a response.

To terminate the authentication process, use the following commands.

- Terminate the authentication process due to an unresponsive supplicant.

  INTERFACE mode

  ```
  dot1x supplicant-timeout seconds
  ```

  The range is from 1 to 300.

  The default is **30**.
- Terminate the authentication process due to an unresponsive authentication server.

  INTERFACE mode

  ```
  dot1x server-timeout seconds
  ```

  The range is from 1 to 300.

  The default is **30**.

Example of Viewing Configured Server Timeouts

The example shows configuration information for a port for which the authenticator terminates the authentication process for an unresponsive supplicant or server after 15 seconds.

The bold lines show the new supplicant and server timeouts.

```
Dell(conf-if-Te-0/0)#dot1x port-control force-authorized
Dell(conf-if-Te-0/0)#do show dot1x interface TenGigabitEthernet 0/0

802.1x information on Te 0/0:
-----------------------------
Dot1x Status:             Enable
Port Control:             FORCE_AUTHORIZED
Port Auth Status:         UNAUTHORIZED
Re-Authentication:        Disable
Untagged VLAN id:         None
Guest VLAN:               Disable
Guest VLAN id:            NONE
Auth-Fail VLAN:           Disable
Auth-Fail VLAN id:        NONE
Auth-Fail Max-Attempts:   NONE
Tx Period:                90 seconds
Quiet Period:             120 seconds
ReAuth Max:               10
Supplicant Timeout:     15 seconds
Server Timeout:       15 seconds
Re-Auth Interval:         7200 seconds
Max-EAP-Req:              10

Auth Type:                SINGLE_HOST
Auth PAE State:           Initialize
Backend State:            Initialize
```

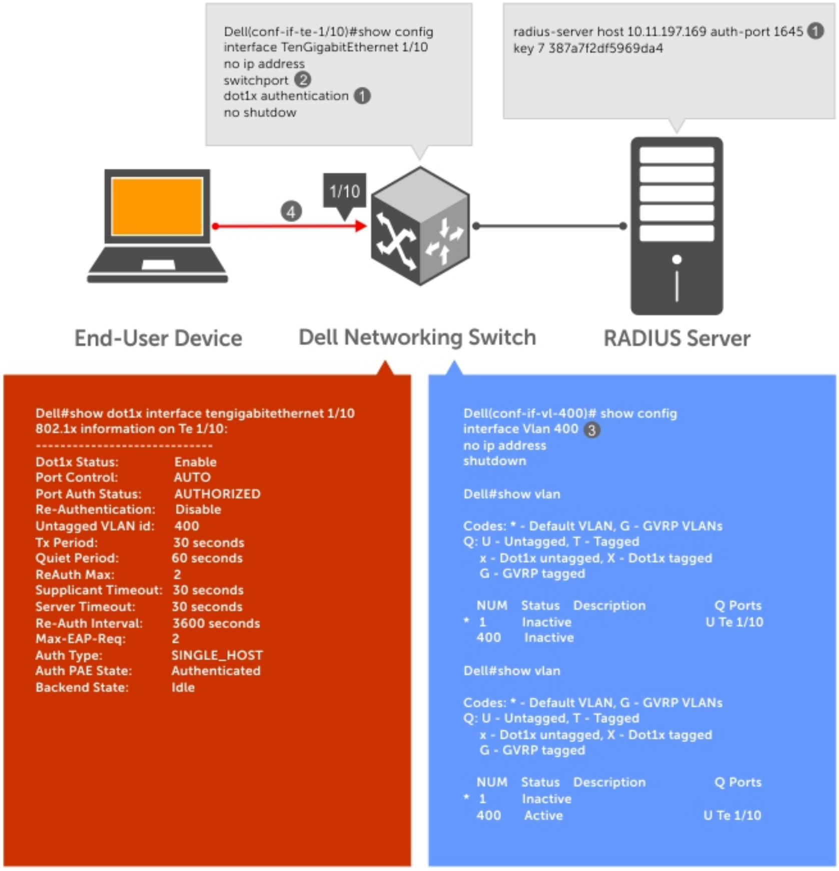
Enter the tasks the user should do after finishing this task (optional).

Configuring Dynamic VLAN Assignment with Port Authentication

Dell Networking OS supports dynamic VLAN assignment when using 802.1X.
The basis for VLAN assignment is RADIUS attribute 81, Tunnel-Private-Group-ID. Dynamic VLAN assignment uses the standard dot1x procedure:

1. The host sends a dot1x packet to the Dell Networking system
2. The system forwards a RADIUS REQEST packet containing the host MAC address and ingress port number
3. The RADIUS server authenticates the request and returns a RADIUS ACCEPT message with the VLAN assignment using Tunnel-Private-Group-ID

The illustration shows the configuration on the Dell Networking system before connecting the end user device in black and blue text, and after connecting the device in red text. The blue text corresponds to the preceding numbered steps on dynamic VLAN assignment with 802.1X.



```
Dell(conf-if-te-1/10)#show config
interface TenGigabitEthernet 1/10
no ip address
switchport ②
dot1x authentication ①
no shutdow
```

```
radius-server host 10.11.197.169 auth-port 1645 ①
key 7 387a7f2df5969da4
```

```
Dell#show dot1x interface tengigabitethernet 1/10
802.1x information on Te 1/10:
-----------------------------
Dot1x Status:           Enable
Port Control:           AUTO
Port Auth Status:       AUTHORIZED
Re-Authentication:      Disable
Untagged VLAN id:       400
Tx Period:              30 seconds
Quiet Period:           60 seconds
ReAuth Max:             2
Supplicant Timeout:     30 seconds
Server Timeout:         30 seconds
Re-Auth Interval:       3600 seconds
Max-EAP-Req:            2
Auth Type:              SINGLE_HOST
Auth PAE State:         Authenticated
Backend State:          Idle
```

```
Dell(conf-if-vl-400)# show config
interface Vlan 400 ③
no ip address
shutdown

Dell#show vlan

Codes: * - Default VLAN, G - GVRP VLANs
Q: U - Untagged, T - Tagged
   x - Dot1x untagged, X - Dot1x tagged
   G - GVRP tagged

  NUM   Status    Description          Q Ports
* 1     Inactive                       U Te 1/10
  400   Inactive

Dell#show vlan

Codes: * - Default VLAN, G - GVRP VLANs
Q: U - Untagged, T - Tagged
   x - Dot1x untagged, X - Dot1x tagged
   G - GVRP tagged

  NUM   Status    Description          Q Ports
* 1     Inactive
  400   Active                         U Te 1/10
```

Figure 7. Dynamic VLAN Assignment

1. Configure 8021.x globally (refer to Enabling 802.1X) along with relevant RADIUS server configurations (refer to the illustration inDynamic VLAN Assignment with Port Authentication).
2. Make the interface a switchport so that it can be assigned to a VLAN.
3. Create the VLAN to which the interface will be assigned.
4. Connect the supplicant to the port configured for 802.1X.
5. Verify that the port has been authorized and placed in the desired VLAN (refer to the illustration in Dynamic VLAN Assignment with Port Authentication).

Guest and Authentication-Fail VLANs

Typically, the authenticator (the Dell system) denies the supplicant access to the network until the supplicant is authenticated. If the supplicant is authenticated, the authenticator enables the port and places it in either the VLAN for which the port is configured or the VLAN that the authentication server indicates in the authentication data.

NOTE: Ports cannot be dynamically assigned to the default VLAN.

If the supplicant fails authentication, the authenticator typically does not enable the port. In some cases this behavior is not appropriate. External users of an enterprise network, for example, might not be able to be authenticated, but still need access to the network. Also, some dumb-terminals, such as network printers, do not have 802.1X capability and therefore cannot authenticate themselves. To be able to connect such devices, they must be allowed access the network without compromising network security.

The Guest VLAN 802.1X extension addresses this limitation with regard to non-802.1X capable devices and the Authentication-fail VLAN 802.1X extension addresses this limitation with regard to external users.

- If the supplicant fails authentication a specified number of times, the authenticator places the port in the Authentication-fail VLAN.
- If a port is already forwarding on the Guest VLAN when 802.1X is enabled, the port is moved out of the Guest VLAN and the authentication process begins.

Configuring a Guest VLAN

If the supplicant does not respond within a determined amount of time ([reauth-max + 1] * tx-period, the system assumes that the host does not have 802.1X capability and the port is placed in the Guest VLAN.

NOTE: For more information about configuring timeouts, refer to Configuring Timeouts.

Configure a port to be placed in the Guest VLAN after failing to respond within the timeout period using the `dot1x guest-vlan` command from INTERFACE mode. View your configuration using the `show config` command from INTERFACE mode or using the `show dot1x interface` command from EXEC Privilege mode.

Example of Viewing Guest VLAN Configuration

```
Dell(conf-if-Te-2/1)#dot1x guest-vlan 200
Dell(conf-if-Te 2/1))#show config
!
interface TenGigabitEthernet 21
  switchport
  dot1x guest-vlan 200
  no shutdown
Dell(conf-if-Te 2/1))#
```

Configuring an Authentication-Fail VLAN

If the supplicant fails authentication, the authenticator re-attempts to authenticate after a specified amount of time.

NOTE: For more information about authenticator re-attempts, refer to Configuring a Quiet Period after a Failed Authentication.

You can configure the maximum number of times the authenticator re-attempts authentication after a failure (**3** by default), after which the port is placed in the Authentication-fail VLAN.

Configure a port to be placed in the VLAN after failing the authentication process as specified number of times using the `dot1x auth-fail-vlan` command from INTERFACE mode. Configure the maximum number of authentication attempts by the authenticator using the keyword `max-attempts` with this command.

Example of Configuring Maximum Authentication Attempts

```
Dell(conf-if-Te-2/1)#dot1x guest-vlan 200
Dell(conf-if-Te 2/1)#show config
```

```
!
interface TenGigabitEthernet 2/1
  switchport
  dot1x authentication
  dot1x guest-vlan 200
no shutdown
Dell(conf-if-Te-2/1)#

Dell(conf-if-Te-2/1)#dot1x auth-fail-vlan 100 max-attempts 5
Dell(conf-if-Te-2/1)#show config
!
interface TenGigabitEthernet 2/1
  switchport
  dot1x authentication
  dot1x guest-vlan 200
  dot1x auth-fail-vlan 100 max-attempts 5
no shutdown
Dell(conf-if-Te-2/1)#
```

Example of Viewing Configured Authentication

View your configuration using the `show config` command from INTERFACE mode, as shown in the example in Configuring a Guest VLAN or using the `show dot1x interface` command from EXEC Privilege mode.

```
802.1x information on Te 2/1:
-----------------------------
Dot1x Status:            Enable
Port Control:            FORCE_AUTHORIZED
Port Auth Status:        UNAUTHORIZED
Re-Authentication:       Disable
Untagged VLAN id:        None
Guest VLAN:          Disabled
Guest VLAN id:        200
Auth-Fail VLAN:        Disabled
Auth-Fail VLAN id:     100
Auth-Fail Max-Attempts: 5
Tx Period:               90 seconds
Quiet Period:            120 seconds
ReAuth Max:              10
Supplicant Timeout:      15 seconds
Server Timeout:          15 seconds
Re-Auth Interval:        7200 seconds
Max-EAP-Req:             10
Auth Type:               SINGLE_HOST

Auth PAE State:          Initialize
Backend State:           Initialize
```

6

Access Control Lists (ACLs)

This chapter describes access control lists (ACLs), prefix lists, and route-maps.

- Access control lists (ACLs), *Ingress* IP and MAC ACLs , and *Egress* IP and MAC ACLs are supported on the S6000 platform.

At their simplest, access control lists (ACLs), prefix lists, and route-maps permit or deny traffic based on MAC and/or IP addresses. This chapter describes implementing IP ACLs, IP prefix lists and route-maps. For MAC ACLS, refer to Layer 2.

An ACL is essentially a filter containing some criteria to match (examine IP, transmission control protocol [TCP], or user datagram protocol [UDP] packets) and an action to take (permit or deny). ACLs are processed in sequence so that if a packet does not match the criterion in the first filter, the second filter (if configured) is applied. When a packet matches a filter, the switch drops or forwards the packet based on the filter's specified action. If the packet does not match any of the filters in the ACL, the packet is dropped (implicit deny).

The number of ACLs supported on a system depends on your content addressable memory (CAM) size. For more information, refer to User Configurable CAM Allocation and CAM Optimization. For complete CAM profiling information, refer to Content Addressable Memory (CAM).

Starting from the release 9.4.(0.0), you can configure ACLs on VRF instances. In addition to the existing qualifying parameters, Layer 3 ACLs also incorporate VRF ID as one of the parameters. Using this new capability, you can also configure VRF based ACLs on interfaces.

NOTE: You can apply Layer 3 VRF-aware ACLs only at the ingress level.

You can apply VRF-aware ACLs on:

- VRF Instances
- Interfaces

In order to configure VRF-aware ACLs on VRF instances, you must carve out a separate CAM region. You can use the `cam-acl` command for allocating CAM regions. As part of the enhancements to support VRF-aware ACLs, the `cam-acl` command now includes the following new parameter that enables you to allocate a CAM region: *vrfv4acl*.

The order of priority for configuring user-defined ACL CAM regions is as follows:

- V4 ACL CAM
- VRF V4 ACL CAM
- L2 ACL CAM

With the inclusion of VRF based ACLs, the order of precedence of Layer 3 ACL rules is as follows:

- Port/VLAN based PERMIT/DENY Rules

- Port/VLAN based IMPLICIT DENY Rules
- VRF based PERMIT/DENY Rules
- VRF based IMPLICIT DENY Rules

NOTE: In order for the VRF ACLs to take effect, ACLs configured in the Layer 3 CAM region must have an implicit-permit option.

You can use the `ip access-group` command to configure VRF-aware ACLs on interfaces. Using the `ip access-group` command, in addition to a range of VLANs, you can also specify a range of VRFs as input for configuring ACLs on interfaces. The VRF range is from 1 to 63. These ACLs use the existing V4 ACL CAM region to populate the entries in the hardware and do not require you to carve out a separate CAM region.

NOTE: You can configure VRF-aware ACLs on interfaces either using a range of VLANs or a range of VRFs but not both.

IP Access Control Lists (ACLs)

In Dell Networking switch/routers, you can create two different types of IP ACLs: standard or extended.

A standard ACL filters packets based on the source IP packet. An extended ACL filters traffic based on the following criteria:

- IP protocol number
- Source IP address
- Destination IP address
- Source TCP port number
- Destination TCP port number
- Source UDP port number
- Destination UDP port number

For more information about ACL options, refer to the *Dell Networking OS Command Reference Guide*.

For extended ACL, TCP, and UDP filters, you can match criteria on specific or ranges of TCP or UDP ports. For extended ACL TCP filters, you can also match criteria on established TCP sessions.

When creating an access list, the sequence of the filters is important. You have a choice of assigning sequence numbers to the filters as you enter them, or the Dell Networking Operating System (OS) assigns numbers in the order the filters are created. The sequence numbers are listed in the display output of the `show config` and `show ip accounting access-list` commands.

Ingress and egress Hot Lock ACLs allow you to append or delete new rules into an existing ACL (already written into CAM) without disrupting traffic flow. Existing entries in the CAM are shuffled to accommodate the new entries. Hot lock ACLs are enabled by default and support both standard and extended ACLs and on all platforms.

NOTE: Hot lock ACLs are supported for Ingress ACLs only.

CAM Usage

The following section describes CAM allocation and CAM optimization.

- User Configurable CAM Allocation

- CAM Optimization

User Configurable CAM Allocation

User configurable CAM allocations are supported on the S6000 platform.

Allocate space for IPV6 ACLs by using the `cam-acl` command in CONFIGURATION mode.

The CAM space is allotted in filter processor (FP) blocks. The total space allocated must equal 13 FP blocks. (There are 16 FP blocks, but System Flow requires three blocks that cannot be reallocated.)

Enter the `ipv6acl` allocation as a factor of 2 (2, 4, 6, 8, 10). All other profile allocations can use either even or odd numbered ranges.

If you want to configure ACL's on VRF instances, you must allocate a CAM region using the *vrfv4acl* option in the **cam-acl** command.

Save the new CAM settings to the startup-config (use `write-mem` or `copy run start`) then reload the system for the new settings to take effect.

CAM Optimization

The `CAM optimization` command is supported on the S6000 platform.

When you enable this command, if a policy map containing classification rules (ACL and/or dscp/ ip-precedence rules) is applied to more than one physical interface on the same port-pipe, only a single copy of the policy is written (only one FP entry is used). When you disable this command, the system behaves as described in this chapter.

Test CAM Usage

The `test cam-usage` command is supported on the S6000 platform.

This command applies to both IPv4 and IPv6 CAM profiles, but is best used when verifying QoS optimization for IPv6 ACLs.

To determine whether sufficient ACL CAM space is available to enable a service-policy, use this command. To verify the actual CAM space required, create a class map with all the required ACL rules, then execute the `test cam-usage` command in Privilege mode. The following example shows the output when executing this command. The status column indicates whether you can enable the policy.

Example of the `test cam-usage` Command

```
Dell#test cam-usage service-policy input TestPolicy linecard all

Linecard|Portpipe|CAM Partition|Available CAM|Estimated CAM per Port|Status
--------------------------------------------------------------------------
       2|       1|     IPv4Flow|          232|                     0|Allowed
       2|       1|     IPv6Flow|            0|                     0|Allowed
       4|       0|     IPv4Flow|          232|                     0|Allowed
       4|       0|     IPv6Flow|            0|                     0|Allowed
Dell#
```

Implementing ACLs on Dell Networking OS

You can assign one IP ACL per interface with Dell Networking OS. If you do not assign an IP ACL to an interface, it is not used by the software in any other capacity.

The number of entries allowed per ACL is hardware-dependent. For detailed specification on entries allowed per ACL, refer to your line card documentation.

If counters are enabled on ACL rules that are already configured, those counters are reset when a new rule which is inserted or prepended or appended requires a hardware shift in the flow table. Resetting the counters to 0 is transient as the proginal counter values are retained after a few seconds. If there is no need to shift the flow in the hardware, the counters are not disturbed. This is applicable to the following features:

- L2 Ingress Access list
- L2 Egress Access list

NOTE: IP ACLs are supported over VLANs in Dell Networking OS version 6.2.1.1 and higher.

ACLs and VLANs

There are some differences when assigning ACLs to a VLAN rather than a physical port.

For example, when using a single port-pipe, if you apply an ACL to a VLAN, one copy of the ACL entries is installed in the ACL CAM on the port-pipe. The entry looks for the incoming VLAN in the packet. Whereas if you apply an ACL on individual ports of a VLAN, separate copies of the ACL entries are installed for each port belonging to a port-pipe.

When you use the `log` keyword, the CP has to log the details about the packets that match. Depending on how many packets match the log entry and at what rate, the CP might become busy as it has to log these packets' details. However, the other processors (RP1 and RP2) are unaffected. This option is typically useful when debugging some problem related to control traffic. We have used this option numerous times in the field and have not encountered problems so far.

ACL Optimization

If an access list contains duplicate entries, Dell Networking OS deletes one entry to conserve CAM space.

Standard and extended ACLs take up the same amount of CAM space. A single ACL rule uses two CAM entries whether it is identified as a standard or extended ACL.

Determine the Order in which ACLs are Used to Classify Traffic

When you link class-maps to queues using the `service-queue` command, Dell Networking OS matches the class-maps according to queue priority (queue numbers closer to 0 have lower priorities).

As shown in the following example, class-map *cmap2* is matched against ingress packets before *cmap1*.

ACLs *acl1* and *acl2* have overlapping rules because the address range 20.1.1.0/24 is within 20.0.0.0/8. Therefore (without the keyword order), packets within the range 20.1.1.0/24 match positive against *cmap1* and are buffered in queue 7, though you intended for these packets to match positive against *cmap2* and be buffered in queue 4.

In cases such as these, where class-maps with overlapping ACL rules are applied to different queues, use the `order` keyword to specify the order in which you want to apply ACL rules. The order can range from 0 to 254. Dell Networking OS writes to the CAM ACL rules with lower-order numbers (order numbers closer to 0) before rules with higher-order numbers so that packets are matched as you intended. By default, all ACL rules have an order of **255**.

Example of the `order` Keyword to Determine ACL Sequence

```
Dell(conf)#ip access-list standard acl1
Dell(config-std-nacl)#permit 20.0.0.0/8
Dell(config-std-nacl)#exit
Dell(conf)#ip access-list standard acl2
Dell(config-std-nacl)#permit 20.1.1.0/24 order 0
Dell(config-std-nacl)#exit
Dell(conf)#class-map match-all cmap1
Dell(conf-class-map)#match ip access-group acl1
Dell(conf-class-map)#exit
Dell(conf)#class-map match-all cmap2
Dell(conf-class-map)#match ip access-group acl2
Dell(conf-class-map)#exit
Dell(conf)#policy-map-input pmap
Dell(conf-policy-map-in)#service-queue 7 class-map cmap1
Dell(conf-policy-map-in)#service-queue 4 class-map cmap2
Dell(conf-policy-map-in)#exit
Dell(conf)#interface te 10/0
Dell(conf-if-te-10/0)#service-policy input pmap
```

Important Points to Remember

- For route-maps with more than one match clause:
  - Two or more match clauses within the same route-map sequence have the *same* match commands (though the values are different), matching a packet against these clauses is a logical OR operation.
  - Two or more match clauses within the same route-map sequence have *different* match commands, matching a packet against these clauses is a logical AND operation.
- If no match is found in a route-map sequence, the process moves to the next route-map sequence until a match is found, or there are no more sequences.
- When a match is found, the packet is forwarded and no more route-map sequences are processed.
  - If a continue clause is included in the route-map sequence, the next or a specified route-map sequence is processed after a match is found.

Configuration Task List for Route Maps

Configure route maps in ROUTE-MAP mode and apply the maps in various commands in ROUTER RIP and ROUTER OSPF modes.

The following list includes the configuration tasks for route maps, as described in the following sections.

- Create a route map (mandatory)
- Configure route map filters (optional)
- Configure a route map for route redistribution (optional)
- Configure a route map for route tagging (optional)

Creating a Route Map

Route maps, ACLs, and prefix lists are similar in composition because all three contain filters, but route map filters do not contain the permit and deny actions found in ACLs and prefix lists.
Route map filters match certain routes and set or specify values.

To create a route map, use the following command.

- Create a route map and assign it a unique name. The optional `permit` and `deny` keywords are the action of the route map.

  CONFIGURATION mode

  ```
  route-map map-name [permit | deny] [sequence-number]
  ```

  The default is **permit**.

  The optional `seq` keyword allows you to assign a sequence number to the route map instance.

Configured Route Map Examples

The default action is **permit** and the default sequence number starts at **10**. When you use the keyword `deny` in configuring a route map, routes that meet the match filters are not redistributed.

To view the configuration, use the `show config` command in ROUTE-MAP mode.

```
Dell(config-route-map)#show config
!
route-map dilling permit 10
Dell(config-route-map)#
```

You can create multiple instances of this route map by using the `sequence number` option to place the route maps in the correct order. Dell Networking OS processes the route maps with the lowest sequence number first. When a configured route map is applied to a command, such as `redistribute`, traffic passes through all instances of that route map until a match is found. The following is an example with two instances of a route map.

The following example shows matching instances of a route-map.

```
Dell#show route-map
route-map zakho, permit, sequence 10
  Match clauses:
  Set clauses:
route-map zakho, permit, sequence 20
  Match clauses:
    interface GigabitEthernet 0/1
  Set clauses:
    tag 35
    level stub-area
Dell#
```

To delete all instances of that route map, use the `no route-map` *`map-name`* command. To delete just one instance, add the sequence number to the command syntax.

```
Dell(conf)#no route-map zakho 10
Dell(conf)#end
Dell#show route-map
route-map zakho, permit, sequence 20
  Match clauses:
    interface GigabitEthernet 0/1
  Set clauses:
    tag 35
    level stub-area
Dell#
```

The following example shows a route map with multiple instances. The `show config` command displays only the configuration of the current route map instance. To view all instances of a specific route map, use the `show route-map` command.

```
Dell#show route-map dilling
route-map dilling, permit, sequence 10
  Match clauses:
  Set clauses:
route-map dilling, permit, sequence 15
  Match clauses:
    interface Loopback 23
  Set clauses:
    tag 3444
Dell#
```

To delete a route map, use the `no route-map` *`map-name`* command in CONFIGURATION mode.

Configure Route Map Filters

Within ROUTE-MAP mode, there are `match` and `set` commands.

- `match` commands search for a certain criterion in the routes.
- `set` commands change the characteristics of routes, either adding something or specifying a level.

When there are multiple `match` commands with the same parameter under one instance of route-map, Dell Networking OS does a match between all of those `match` commands. If there are multiple `match` commands with different parameters, Dell Networking OS does a match ONLY if there is a match among ALL the `match` commands.

In the following example, there is a match if a route has any of the tag values specified in the `match` commands.

Example of the `match` Command to Match Any of Several Values

The following example shows using the `match` command to match any of several values.

```
Dell(conf)#route-map force permit 10
Dell(config-route-map)#match tag 1000
Dell(config-route-map)#match tag 2000
Dell(config-route-map)#match tag 3000
```

Example of the `match` Command to Match All Specified Values

In the next example, there is a match *only* if a route has *both* of the specified characteristics. In this example, there a match only if the route has a tag value of 1000 *and* a metric value of 2000.

Also, if there are different instances of the same route-map, then it’s sufficient if a permit match happens in any instance of that route-map.

```
Dell(conf)#route-map force permit 10
Dell(config-route-map)#match tag 1000
Dell(config-route-map)#match metric 2000
```

In the following example, instance 10 permits the route having a tag value of 1000 and instances 20 and 30 deny the route having a tag value of 1000. In this scenario, Dell Networking OS scans all the instances of the route-map for any permit statement. If there is a match anywhere, the route is permitted. However, other instances of the route-map deny it.

Example of the `match` Command to Permit and Deny Routes

```
Dell(conf)#route-map force permit 10
Dell(config-route-map)#match tag 1000

Dell(conf)#route-map force deny 20
Dell(config-route-map)#match tag 1000

Dell(conf)#route-map force deny 30
Dell(config-route-map)#match tag 1000
```

Configuring Match Routes

To configure match criterion for a route map, use the following commands.

- Match routes with the same AS-PATH numbers.

  CONFIG-ROUTE-MAP mode

  ```
  match as-path as-path-name
  ```
- Match routes with COMMUNITY list attributes in their path.

  CONFIG-ROUTE-MAP mode

  ```
  match community community-list-name [exact]
  ```
- Match routes whose next hop is a specific interface.

  CONFIG-ROUTE-MAP mode

  ```
  match interface interface
  ```

  The parameters are:

  - For a Fast Ethernet interface, enter the keyword `FastEthernet` then the slot/ port information.
  - For a 1-Gigabit Ethernet interface, enter the keyword `gigabitEthernet` then the slot/port information.
  - For a loopback interface, enter the keyword `loopback` then a number between zero (0) and 16383.
  - For a port channel interface, enter the keywords `port-channel` then a number.
  - For a SONET interface, enter the keyword `sonet` then the slot/port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `tengigabitEthernet` then the slot/port information.
  - For a VLAN, enter the keyword `vlan` then a number from 1 to 4094.
  - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.
- Match destination routes specified in a prefix list (IPv4).

  CONFIG-ROUTE-MAP mode

  ```
  match ip address prefix-list-name
  ```
- Match destination routes specified in a prefix list (IPv6).

  CONFIG-ROUTE-MAP mode

  ```
  match ipv6 address prefix-list-name
  ```
- Match next-hop routes specified in a prefix list (IPv4).

  CONFIG-ROUTE-MAP mode

  ```
  match ip next-hop {access-list-name | prefix-list prefix-list-name}
  ```

- Match next-hop routes specified in a prefix list (IPv6).

  CONFIG-ROUTE-MAP mode

  ```
  match ipv6 next-hop {access-list-name | prefix-list prefix-list-name}
  ```

- Match source routes specified in a prefix list (IPv4).

  CONFIG-ROUTE-MAP mode

  ```
  match ip route-source {access-list-name | prefix-list prefix-list-name}
  ```

- Match source routes specified in a prefix list (IPv6).

  CONFIG-ROUTE-MAP mode

  ```
  match ipv6 route-source {access-list-name | prefix-list prefix-list-name}
  ```

- Match routes with a specific value.

  CONFIG-ROUTE-MAP mode

  ```
  match metric metric-value
  ```

- Match BGP routes based on the ORIGIN attribute.

  CONFIG-ROUTE-MAP mode

  ```
  match origin {egp | igp | incomplete}
  ```

- Match routes specified as internal or external to OSPF, ISIS level-1, ISIS level-2, or locally generated.

  CONFIG-ROUTE-MAP mode

  ```
  match route-type {external [type-1 | type-2] | internal | level-1 | level-2 |
  local }
  ```

- Match routes with a specific tag.

  CONFIG-ROUTE-MAP mode

  ```
  match tag tag-value
  ```

To create route map instances, use these commands. There is no limit to the number of `match` commands per route map, but the convention is to keep the number of match filters in a route map low. `Set` commands do not require a corresponding `match` command.

Configuring Set Conditions

To configure a set condition, use the following commands.

- Add an AS-PATH number to the beginning of the AS-PATH.

  CONFIG-ROUTE-MAP mode

  ```
  set as-path prepend as-number [... as-number]
  ```

- Generate a tag to be added to redistributed routes.

  CONFIG-ROUTE-MAP mode

  ```
  set automatic-tag
  ```

- Specify an OSPF area or ISIS level for redistributed routes.

  CONFIG-ROUTE-MAP mode

  ```
  set level {backbone | level-1 | level-1-2 | level-2 | stub-area}
  ```

- Specify a value for the BGP route's LOCAL_PREF attribute.

CONFIG-ROUTE-MAP mode

```
set local-preference value
```

- Specify a value for redistributed routes.

  CONFIG-ROUTE-MAP mode

  ```
  set metric {+ | - | metric-value}
  ```

- Specify an OSPF or ISIS type for redistributed routes.

  CONFIG-ROUTE-MAP mode

  ```
  set metric-type {external | internal | type-1 | type-2}
  ```

- Assign an IP address as the route's next hop.

  CONFIG-ROUTE-MAP mode

  ```
  set next-hop ip-address
  ```

- Assign an IPv6 address as the route's next hop.

  CONFIG-ROUTE-MAP mode

  ```
  set ipv6 next-hop ip-address
  ```

- Assign an ORIGIN attribute.

  CONFIG-ROUTE-MAP mode

  ```
  set origin {egp | igp | incomplete}
  ```

- Specify a tag for the redistributed routes.

  CONFIG-ROUTE-MAP mode

  ```
  set tag tag-value
  ```

- Specify a value as the route's weight.

  CONFIG-ROUTE-MAP mode

  ```
  set weight value
  ```

To create route map instances, use these commands. There is no limit to the number of `set` commands per route map, but the convention is to keep the number of set filters in a route map low. `Set` commands do not require a corresponding `match` command.

Configure a Route Map for Route Redistribution

Route maps on their own cannot affect traffic and must be included in different commands to affect routing traffic.

Route redistribution occurs when Dell Networking OS learns the advertising routes from static or directly connected routes or another routing protocol. Different protocols assign different values to redistributed routes to identify either the routes and their origins. The metric value is the most common attribute that is changed to properly redistribute other routes into a routing protocol. Other attributes that can be changed include the metric type (for example, external and internal route types in OSPF) and route tag. Use the `redistribute` command in OSPF, RIP, ISIS, and BGP to set some of these attributes for routes that are redistributed into those protocols.

Route maps add to that redistribution capability by allowing you to match specific routes and set or change more attributes when redistributing those routes.

In the following example, the `redistribute` command calls the route map `static ospf` to redistribute only certain static routes into OSPF. According to the route map `static ospf`, only routes that have a next hop of Gigabitethernet interface 0/0 and that have a metric of 255 are redistributed into the OSPF backbone area.

NOTE: When re-distributing routes using route-maps, you must create the route-map defined in the `redistribute` command under the routing protocol. If you do not create a route-map, NO routes are redistributed.

Example of Calling a Route Map to Redistribute Specified Routes

```
router ospf 34
  default-information originate metric-type 1
  redistribute static metric 20 metric-type 2 tag 0 route-map staticospf
!
route-map staticospf permit 10
  match interface GigabitEthernet 0/0
  match metric 255
  set level backbone
```

Configure a Route Map for Route Tagging

One method for identifying routes from different routing protocols is to assign a tag to routes from that protocol.

As the route enters a different routing domain, it is tagged. The tag is passed along with the route as it passes through different routing protocols. You can use this tag when the route leaves a routing domain to redistribute those routes again.

In the following example, the `redistribute ospf` command with a route map is used in ROUTER RIP mode to apply a tag of 34 to all internal OSPF routes that are redistributed into RIP.

Example of the `redistribute` Command Using a Route Tag

```
!
router rip
  redistribute ospf 34 metric 1 route-map torip
!
route-map torip permit 10
  match route-type internal
  set tag 34
!
```

Continue Clause

Normally, when a match is found, set clauses are executed, and the packet is then forwarded; no more route-map modules are processed.

If you configure the `continue` command at the end of a module, the next module (or a specified module) is processed even after a match is found. The following example shows a continue clause at the end of a route-map module. In this example, if a match is found in the route-map “test” module 10, module 30 is processed.

NOTE: If you configure the continue clause without specifying a module, the next sequential module is processed.

Example of Using the `continue` Clause in a Route Map

```
!
route-map test permit 10
match commu comm-list1
set community 1:1 1:2 1:3
set as-path prepend 1 2 3 4 5
continue 30!
```

IP Fragment Handling

Dell Networking OS supports a configurable option to explicitly deny IP fragmented packets, particularly second and subsequent packets.

It extends the existing ACL command syntax with the `fragments` keyword for all Layer 3 rules applicable to all Layer protocols (permit/deny ip/tcp/udp/icmp).

- Both standard and extended ACLs support IP fragments.
- Second and subsequent fragments are allowed because a Layer 4 rule cannot be applied to these fragments. If the packet is to be denied eventually, the first fragment would be denied and hence the packet as a whole cannot be reassembled.
- Implementing the required rules uses a significant number of CAM entries per TCP/UDP entry.
- For IP ACL, Dell Networking OS always applies implicit deny. You do not have to configure it.
- For IP ACL, Dell Networking OS applies implicit permit for second and subsequent fragment just prior to the implicit deny.
- If you configure an *explicit* deny, the second and subsequent fragments do not hit the implicit permit rule for fragments.
- Loopback interfaces do not support ACLs using the `IP fragment` option. If you configure an ACL with the `fragments` option and apply it to a Loopback interface, the command is accepted but the ACL entries are not actually installed the offending rule in CAM.

IP Fragments ACL Examples

The following examples show how you can use ACL commands with the `fragment` keyword to filter fragmented packets.

Example of Permitting All Packets on an Interface

The following configuration permits all packets (both fragmented and non-fragmented) with destination IP 10.1.1.1. The second rule does not get hit at all.

```
Dell(conf)#ip access-list extended ABC
Dell(conf-ext-nacl)#permit ip any 10.1.1.1/32FTOS(conf-ext-nacl)#deny ip any
10.1.1.1./32 fragments
Dell(conf-ext-nacl)
```

Example of Denying Second and Subsequent Fragments

To deny the second/subsequent fragments, use the same rules in a different order. These ACLs deny all second and subsequent fragments with destination IP 10.1.1.1 but permit the first fragment and non-fragmented packets with destination IP 10.1.1.1.

```
Dell(conf)#ip access-list extended ABC
Dell(conf-ext-nacl)#deny ip any 10.1.1.1/32 fragments
Dell(conf-ext-nacl)#permit ip any 10.1.1.1/32
Dell(conf-ext-nacl)
```

Layer 4 ACL Rules Examples

The following examples show the ACL commands for Layer 4 packet filtering.

Permit an ACL line with L3 information only, and the `fragments` keyword is present:

If a packet’s L3 information matches the L3 information in the ACL line, the packet's FO is checked.

- If a packet's FO > 0, the packet is permitted.
- If a packet's FO = 0, the next ACL entry is processed.

Deny ACL line with L3 information only, and the `fragments` keyword is present:

If a packet's L3 information does match the L3 information in the ACL line, the packet's FO is checked.

- If a packet's FO > 0, the packet is denied.
- If a packet's FO = 0, the next ACL line is processed.

Example of Permitting All Packets from a Specified Host

In this first example, TCP packets from host 10.1.1.1 with TCP destination port equal to 24 are permitted. All others are denied.

```
Dell(conf)#ip access-list extended ABC
Dell(conf-ext-nacl)#permit tcp host 10.1.1.1 any eq 24
Dell(conf-ext-nacl)#deny ip any any fragment
Dell(conf-ext-nacl)
```

Example of Permitting Only First Fragments and Non-Fragmented Packets from a Specified Host

In the following example, the TCP packets that are first fragments or non-fragmented from host 10.1.1.1 with TCP destination port equal to 24 are permitted. Additionally, all TCP non-first fragments from host 10.1.1.1 are permitted. All other IP packets that are non-first fragments are denied.

```
Dell(conf)#ip access-list extended ABC
Dell(conf-ext-nacl)#permit tcp host 10.1.1.1 any eq 24
Dell(conf-ext-nacl)#permit tcp host 10.1.1.1 any fragment
Dell(conf-ext-nacl)#deny ip any any fragment
Dell(conf-ext-nacl)
```

Example of Logging Denied Packets

To log all the packets denied and to override the implicit deny rule and the implicit permit rule for TCP/UDP fragments, use a configuration similar to the following.

```
Dell(conf)#ip access-list extended ABC
Dell(conf-ext-nacl)#permit tcp any any fragment
Dell(conf-ext-nacl)#permit udp any any fragment
Dell(conf-ext-nacl)#deny ip any any log
Dell(conf-ext-nacl)
```

When configuring ACLs with the fragments keyword, be aware of the following.

When an ACL filters packets, it looks at the fragment offset (FO) to determine whether it is a fragment.

- FO = 0 means it is either the first fragment or the packet is a non-fragment.
- FO > 0 means it is dealing with the fragments of the original packet.

Configure a Standard IP ACL

To configure an ACL, use commands in IP ACCESS LIST mode and INTERFACE mode.
For a complete list of all the commands related to IP ACLs, refer to the *Dell Networking OS Command Line Interface Reference Guide*. To set up extended ACLs, refer to Configure an Extended IP ACL.
A standard IP ACL uses the source IP address as its match criterion.

1. Enter IP ACCESS LIST mode by naming a standard IP access list.

   CONFIGURATION mode

   ```
   ip access-list standard access-listname
   ```

2. Configure a drop or forward filter.

   CONFIG-STD-NACL mode

   ```
   seq sequence-number {deny | permit} {source [mask] | any | host ip-address}
   [count [byte] [dscp] [order] [fragments]
   ```

NOTE: When assigning sequence numbers to filters, keep in mind that you might need to insert a new filter. To prevent reconfiguring multiple filters, assign sequence numbers in multiples of five.

To view the rules of a particular ACL configured on a particular interface, use the `show ip accounting access-list` *`ACL-name`* `interface` *`interface`* command in EXEC Privilege mode.

Example of Viewing the Rules of a Specific ACL on an Interface

The following is an example of viewing the rules of a specific ACL on an interface.

```
Dell#show ip accounting access-list ToOspf interface gig 1/6
Standard IP access list ToOspf
  seq 5 deny any
  seq 10 deny 10.2.0.0 /16
  seq 15 deny 10.3.0.0 /16
  seq 20 deny 10.4.0.0 /16
  seq 25 deny 10.5.0.0 /16
  seq 30 deny 10.6.0.0 /16
  seq 35 deny 10.7.0.0 /16
  seq 40 deny 10.8.0.0 /16
  seq 45 deny 10.9.0.0 /16
  seq 50 deny 10.10.0.0 /16
Dell#
```

The following example shows how the `seq` command orders the filters according to the sequence number assigned. In the example, filter 25 was configured before filter 15, but the `show config` command displays the filters in the correct order.

```
Dell(config-std-nacl)#seq 25 deny ip host 10.5.0.0 any log
Dell(config-std-nacl)#seq 15 permit tcp 10.3.0.0 /16 any
Dell(config-std-nacl)#show config
!
ip access-list standard dilling
  seq 15 permit tcp 10.3.0.0/16 any
  seq 25 deny ip host 10.5.0.0 any log
Dell(config-std-nacl)#
```

To delete a filter, use the `no seq sequence-number` command in IP ACCESS LIST mode.

If you are creating a standard ACL with only one or two filters, you can let Dell Networking OS assign a sequence number based on the order in which the filters are configured. The software assigns filters in multiples of 5.

Configuring a Standard IP ACL Filter

If you are creating a standard ACL with only one or two filters, you can let Dell Networking OS assign a sequence number based on the order in which the filters are configured. The software assigns filters in multiples of five.

1. Configure a standard IP ACL and assign it a unique name.

   CONFIGURATION mode

   ```
   ip access-list standard access-list-name
   ```
2. Configure a drop or forward IP ACL filter.

   CONFIG-STD-NACL mode

   ```
   {deny | permit} {source [mask] | any | host ip-address} [count [byte] [dscp]
   [order] [fragments]
   ```

When you use the `log` keyword, the CP logs details about the packets that match. Depending on how many packets match the log entry and at what rate, the CP may become busy as it has to log these packets' details.

The following example shows a standard IP ACL in which Dell Networking OS assigns the sequence numbers. The filters were assigned sequence numbers based on the order in which they were configured (for example, the first filter was given the lowest sequence number). The `show config` command in IP ACCESS LIST mode displays the two filters with the sequence numbers 5 and 10.

Example of Viewing a Filter Sequence for a Specified Standard ACL and for an Interface

```
Dell(config-route-map)#ip access standard kigali
Dell(config-std-nacl)#permit 10.1.0.0/16
Dell(config-std-nacl)#show config
!
ip access-list standard kigali
  seq 5 permit 10.1.0.0/16  seq 10 deny tcp any any eq 111
Dell(config-std-nacl)#
```

To view all configured IP ACLs, use the `show ip accounting access-list` command in EXEC Privilege mode.

The following examples shows how to view a standard ACL filter sequence for an interface.

```
Dell#show ip accounting access example interface gig 4/12
Extended IP access list example
  seq 15 deny udp any any eq 111
  seq 20 deny udp any any eq 2049
  seq 25 deny udp any any eq 31337
  seq 30 deny tcp any any range 12345 12346
  seq 35 permit udp host 10.21.126.225 10.4.5.0 /28
  seq 40 permit udp host 10.21.126.226 10.4.5.0 /28
  seq 45 permit udp 10.8.0.0 /16 10.50.188.118 /31 range 1812 1813
  seq 50 permit tcp 10.8.0.0 /16 10.50.188.118 /31 eq 49
  seq 55 permit udp 10.15.1.0 /24 10.50.188.118 /31 range 1812 1813
```

To delete a filter, enter the `show config` command in IP ACCESS LIST mode and locate the sequence number of the filter you want to delete. Then use the `no seq sequence-number` command in IP ACCESS LIST mode.

Configure an Extended IP ACL

Extended IP ACLs filter on source and destination IP addresses, IP host addresses, TCP addresses, TCP host addresses, UDP addresses, and UDP host addresses.

Because traffic passes through the filter in the order of the filter's sequence, you can configure the extended IP ACL by first entering IP ACCESS LIST mode and then assigning a sequence number to the filter.

Configuring Filters with a Sequence Number

To configure filters with a sequence number, use the following commands.

1. Enter IP ACCESS LIST mode by creating an extended IP ACL.
   CONFIGURATION mode

   ```
   ip access-list extended access-list-name
   ```
2. Configure a drop or forward filter.
   CONFIG-EXT-NACL mode

   ```
   seq sequence-number {deny | permit} {ip-protocol-number | icmp | ip | tcp |
   udp} {source mask | any | host ip-address} {destination mask | any | host
   ip-address} [operator port [port]] [count [byte]] [order] [fragments]
   ```

When you use the `log` keyword, the CP logs details about the packets that match. Depending on how many packets match the log entry and at what rate, the CP may become busy as it has to log these packets' details.

Configure Filters, TCP Packets

To create a filter for TCP packets with a specified sequence number, use the following commands.

1. Create an extended IP ACL and assign it a unique name.
   CONFIGURATION mode

   ```
   ip access-list extended access-list-name
   ```
2. Configure an extended IP ACL filter for TCP packets.
   CONFIG-EXT-NACL mode

   ```
   seq sequence-number {deny | permit} tcp {source mask | any | host ip-
   address}} [count [byte]] [order] [fragments]
   ```

Configure Filters, TCP Packets

To create a filter for UDP packets with a specified sequence number, use the following commands.

1. Create an extended IP ACL and assign it a unique name.
   CONFIGURATION mode

`ip access-list extended` *`access-list-name`*

2. Configure an extended IP ACL filter for UDP packets.

   CONFIG-EXT-NACL mode

   ```
   seq sequence-number {deny | permit} tcp {source mask | any | host ip-
   address}} [count [byte]] [order] [fragments]
   ```

Example of the `seq` Command

When you create the filters with a specific sequence number, you can create the filters in any order and the filters are placed in the correct order.

NOTE: When assigning sequence numbers to filters, you may have to insert a new filter. To prevent reconfiguring multiple filters, assign sequence numbers in multiples of five or another number.

The example below shows how the `seq` command orders the filters according to the sequence number assigned. In the example, filter 15 was configured before filter 5, but the `show config` command displays the filters in the correct order.

```
Dell(config-ext-nacl)#seq 15 deny ip host 112.45.0.0 any log
Dell(config-ext-nacl)#seq 5 permit tcp 12.1.3.45 0.0.255.255 any
Dell(config-ext-nacl)#show confi
!
ip access-list extended dilling
  seq 5 permit tcp 12.1.0.0 0.0.255.255 any
  seq 15 deny ip host 112.45.0.0 any log
Dell(config-ext-nacl)#
```

Configuring Filters Without a Sequence Number

If you are creating an extended ACL with only one or two filters, you can let Dell Networking OS assign a sequence number based on the order in which the filters are configured. Dell Networking OS assigns filters in multiples of five.
To configure a filter for an extended IP ACL without a specified sequence number, use any or all of the following commands:

- Configure a deny or permit filter to examine IP packets.

  CONFIG-EXT-NACL mode

  ```
  {deny | permit} {source mask | any | host ip-address} [count [byte]] [order]
  [fragments]
  ```
- Configure a deny or permit filter to examine TCP packets.

  CONFIG-EXT-NACL mode

  ```
  {deny | permit} tcp {source mask] | any | host ip-address}} [count [byte]]
  [order] [fragments]
  ```
- Configure a deny or permit filter to examine UDP packets.

  CONFIG-EXT-NACL mode

  ```
  {deny | permit} udp {source mask | any | host ip-address}} [count [byte]]
  [order] [fragments]
  ```

When you use the `log` keyword, the CP logs details about the packets that match. Depending on how many packets match the log entry and at what rate, the CP may become busy as it has to log these packets' details.

The following example shows an extended IP ACL in which the sequence numbers were assigned by the software. The filters were assigned sequence numbers based on the order in which they were configured (for example, the first filter was given the lowest sequence number). The `show config` command in IP ACCESS LIST mode displays the two filters with the sequence numbers 5 and 10.

Example of Viewing Filter Sequence for a Specified Extended ACL

```
Dell(config-ext-nacl)#deny tcp host 123.55.34.0 any
Dell(config-ext-nacl)#permit udp 154.44.123.34 0.0.255.255 host 34.6.0.0
Dell(config-ext-nacl)#show config
!
ip access-list extended nimule
  seq 5 deny tcp host 123.55.34.0 any
  seq 10 permit udp 154.44.0.0 0.0.255.255 host 34.6.0.0
Dell(config-ext-nacl)#
```

To view all configured IP ACLs and the number of packets processed through the ACL, use the `show ip accounting access-list` command in EXEC Privilege mode, as shown in the first example in Configure a Standard IP ACL Filter.

Configure Layer 2 and Layer 3 ACLs

Both Layer 2 and Layer 3 ACLs may be configured on an interface in Layer 2 mode.

If both L2 and L3 ACLs are applied to an interface, the following rules apply:

- When Dell Networking OS routes the packets, only the L3 ACL governs them because they are not filtered against an L2 ACL.
- When Dell Networking OS switches the packets, first the L3 ACL filters them, then the L2 ACL filters them.
- When Dell Networking OS switches the packets, the egress L3 ACL does not filter the packet.

For the following features, if you enable counters on rules that have already been configured and a new rule is either inserted or prepended, all the existing counters are reset:

- L2 ingress access list
- L3 egress access list
- L2 egress access list

If a rule is simply appended, existing counters are not affected.

Table 4. L2 and L3 Filtering on Switched Packets

L2 ACL Behavior	L3 ACL Behavior	Decision on Targeted Traffic
Deny	Deny	L3 ACL denies.
Deny	Permit	L3 ACL permits.
Permit	Deny	L3 ACL denies.
Permit	Permit	L3 ACL permits.

NOTE: If you configure an interface as a vlan-stack access port, only the L2 ACL filters the packets. The L3 ACL applied to such a port does not affect traffic. That is, existing rules for other features (such as trace-list, policy-based routing [PBR], and QoS) are applied to the permitted traffic.

For information about MAC ACLs, refer to Layer 2.

Assign an IP ACL to an Interface

To pass traffic through a configured IP ACL, assign that ACL to a physical interface, a port channel interface, or a VLAN.

The IP ACL is applied to all traffic entering a physical or port channel interface and the traffic is either forwarded or dropped depending on the criteria and actions specified in the ACL.

The same ACL may be applied to different interfaces and that changes its functionality. For example, you can take ACL "ABCD" and apply it using the `in` keyword and it becomes an ingress access list. If you apply the same ACL using the `out` keyword, it becomes an egress access list. If you apply the same ACL to the Loopback interface, it becomes a Loopback access list.

This section describes the following:

- Configure Ingress ACLs
- Configure Egress ACLs

For more information about Layer-3 interfaces, refer to Interfaces.

Applying an IP ACL

To apply an IP ACL (standard or extended) to a physical or port channel interface, use the following commands.

1. Enter the interface number.

   CONFIGURATION mode

   ```
   interface interface slot/port
   ```

2. Configure an IP address for the interface, placing it in Layer-3 mode.

   INTERFACE mode

   ```
   ip address ip-address
   ```

3. Apply an IP ACL to traffic entering or exiting an interface.

   INTERFACE mode

   ```
   ip access-group access-list-name {in} [implicit-permit] [vlan vlan-range |
   vrf vrf-range]
   ```

   NOTE: The number of entries allowed per ACL is hardware-dependent. For detailed specification about entries allowed per ACL, refer to your line card documentation.

4. Apply rules to the new ACL.

   INTERFACE mode

   ```
   ip access-list [standard | extended] name
   ```

To view which IP ACL is applied to an interface, use the `show config` command in INTERFACE mode, or use the `show running-config` command in EXEC mode.

Example of Viewing ACLs Applied to an Interface

```
Dell(conf-if)#show conf
!
interface GigabitEthernet 0/0
  ip address 10.2.1.100 255.255.255.0
  ip access-group nimule in
  no shutdown
Dell(conf-if)#
```

To filter traffic on Telnet sessions, use only standard ACLs in the `access-class` command.

Counting ACL Hits

You can view the number of packets matching the ACL by using the `count` option when creating ACL entries.

1. Create an ACL that uses rules with the count option. Refer to Configure a Standard IP ACL Filter.
2. Apply the ACL as an inbound or outbound ACL on an interface. Refer to Assign an IP ACL to an Interface.
3. `show ip accounting access-list`

   EXEC Privilege mode

   View the number of packets matching the ACL.

Configure Ingress ACLs

Ingress ACLs are applied to interfaces and to traffic entering the system.

These system-wide ACLs eliminate the need to apply ACLs onto each interface and achieves the same results. By localizing target traffic, it is a simpler implementation.

To create an ingress ACL, use the `ip access-group` command in EXEC Privilege mode. The example shows applying the ACL, rules to the newly created access group, and viewing the access list.

Example of Applying ACL Rules to Ingress Traffic and Viewing ACL Configuration

To specify ingress, use the `in` keyword. Begin applying rules to the ACL with the `ip access-list extended` *`abcd`* command. To view the access-list, use the `show` command.

```
Dell(conf)#interface gige 0/0
Dell(conf-if-gige0/0)#ip access-group abcd in
Dell(conf-if-gige0/0)#show config
!
gigethernet 0/0
  no ip address
  ip access-group abcd in
  no shutdown
Dell(conf-if-gige0/0)#end
Dell#configure terminal
Dell(conf)#ip access-list extended abcd
Dell(config-ext-nacl)#permit tcp any any
Dell(config-ext-nacl)#deny icmp any any
Dell(config-ext-nacl)#permit 1.1.1.2
Dell(config-ext-nacl)#end
Dell#show ip accounting access-list
!
Extended Ingress IP access list abcd on gigethernet 0/0
```

```
  seq 5 permit tcp any any
  seq 10 deny icmp any any
seq 15 permit 1.1.1.2
```

Configure Egress ACLs

Egress ACLs are supported on the S6000 platform.

Egress ACLs are applied to line cards and affect the traffic leaving the system. Configuring egress ACLs onto physical interfaces protects the system infrastructure from attack — malicious and incidental — by explicitly allowing only authorized traffic. These system-wide ACLs eliminate the need to apply ACLs onto each interface and achieves the same results. By localizing target traffic, it is a simpler implementation.

To restrict egress traffic, use an egress ACL. For example, when a denial of service (DOS) attack traffic is isolated to a specific interface, you can apply an egress ACL to block the flow from the exiting the box, thus protecting downstream devices.

To create an egress ACL, use the `ip access-group` command in EXEC Privilege mode. The example shows viewing the configuration, applying rules to the newly created access group, and viewing the access list.

NOTE: VRF based ACL configurations are not supported on the egress traffic.

Example of Applying ACL Rules to Egress Traffic and Viewing ACL Configuration

To specify ingress, use the `out` keyword. Begin applying rules to the ACL with the `ip access-list extended` *`abcd`* command. To view the access-list, use the `show` command.

```
Dell(conf)#interface gige 0/0
Dell(conf-if-gige0/0)#ip access-group abcd out
Dell(conf-if-gige0/0)#show config
!
gigethernet 0/0
  no ip address
  ip access-group abcd out
  no shutdown
Dell(conf-if-gige0/0)#end
Dell#configure terminal
Dell(conf)#ip access-list extended abcd
Dell(config-ext-nacl)#permit tcp any any
Dell(config-ext-nacl)#deny icmp any any
Dell(config-ext-nacl)#permit 1.1.1.2
Dell(config-ext-nacl)#end
Dell#show ip accounting access-list
!
Extended Ingress IP access list abcd on gigethernet 0/0
  seq 5 permit tcp any any
  seq 10 deny icmp any any
seq 15 permit 1.1.1.2

Dell#configure terminal
Dell(conf)#interface te 0/0
Dell(conf-if-te-0/0)#ip vrf forwarding blue
Dell(conf-if-te-0/0)#show config
!
interface TenGigabitEthernet 0/0
ip vrf forwarding blue
no ip address
shutdown
```

```
Dell(conf-if-te-0/0)#
Dell(conf-if-te-0/0)#
Dell(conf-if-te-0/0)#end
Dell#
```

Applying Egress Layer 3 ACLs (Control-Plane)

By default, packets originated from the system are not filtered by egress ACLs.
For example, if you initiate a ping session from the system and apply an egress ACL to block this type of traffic on the interface, the ACL does not affect that ping traffic. The Control Plane Egress Layer 3 ACL feature enhances IP reachability debugging by implementing control-plane ACLs for CPU-generated and CPU-forwarded traffic. Using permit rules with the `count` option, you can track on a per-flow basis whether CPU-generated and CPU-forwarded packets were transmitted successfully.

1. Apply Egress ACLs to IPv4 system traffic.

   CONFIGURATION mode

   ```
   ip control-plane [egress filter]
   ```
2. Apply Egress ACLs to IPv6 system traffic.

   CONFIGURATION mode

   ```
   ipv6 control-plane [egress filter]
   ```
3. Create a Layer 3 ACL using permit rules with the `count` option to describe the desired CPU traffic.

   CONFIG-NACL mode

   ```
   permit ip {source mask | any | host ip-address} {destination mask | any |
   host ip-address} count
   ```

FTOS Behavior: Virtual router redundancy protocol (VRRP) hellos and internet group management protocol (IGMP) packets are not affected when you enable egress ACL filtering for CPU traffic. Packets sent by the CPU with the source address as the VRRP virtual IP address have the interface MAC address instead of VRRP virtual MAC address.

IP Prefix Lists

Prefix lists are supported on the S6000 platform.

IP prefix lists control routing policy. An IP prefix list is a series of sequential filters that contain a matching criterion (examine IP route prefix) and an action (permit or deny) to process routes. The filters are processed in sequence so that if a route prefix does not match the criterion in the first filter, the second filter (if configured) is applied. When the route prefix matches a filter, Dell Networking OS drops or forwards the packet based on the filter's designated action. If the route prefix does not match any of the filters in the prefix list, the route is dropped (that is, implicit deny).

A route prefix is an IP address pattern that matches on bits within the IP address. The format of a route prefix is A.B.C.D/X where A.B.C.D is a dotted-decimal address and /X is the number of bits that should be matched of the dotted decimal address. For example, in 112.24.0.0/16, the first 16 bits of the address 112.24.0.0 match all addresses between 112.24.0.0 to 112.24.255.255.

The following examples show permit or deny filters for specific routes using the `le` and `ge` parameters, where x.x.x.x/x represents a route prefix:

- To deny only /8 prefixes, enter `deny x.x.x.x/x ge 8 le 8.`

- To permit routes with the mask greater than /8 but less than /12, enter `permit x.x.x.x/x ge 8`.
- To deny routes with a mask less than /24, enter `deny x.x.x.x/x le 24`.
- To permit routes with a mask greater than /20, enter `permit x.x.x.x/x ge 20`.

The following rules apply to prefix lists:

- A prefix list without any permit or deny filters allows all routes.
- An "implicit deny" is assumed (that is, the route is dropped) for all route prefixes that do not match a permit or deny filter in a configured prefix list.
- After a route matches a filter, the filter's action is applied. No additional filters are applied to the route.

Implementation Information

In Dell Networking OS, prefix lists are used in processing routes for routing protocols (for example, router information protocol [RIP], open shortest path first [OSPF], and border gateway protocol [BGP]).

NOTE: It is important to know which protocol your system supports prior to implementing prefix-lists.

Configuration Task List for Prefix Lists

To configure a prefix list, use commands in PREFIX LIST, ROUTER RIP, ROUTER OSPF, and ROUTER BGP modes.

Create the prefix list in PREFIX LIST mode and assign that list to commands in ROUTER RIP, ROUTER OSPF and ROUTER BGP modes.

The following list includes the configuration tasks for prefix lists, as described in the following sections.

- Configuring a prefix list
- Use a prefix list for route redistribution

For a complete listing of all commands related to prefix lists, refer to the *Dell Networking OS Command Line Interface Reference Guide*.

Creating a Prefix List

To create a prefix list, use the following commands.

1. Create a prefix list and assign it a unique name.
   You are in PREFIX LIST mode.

   CONFIGURATION mode

   ```
   ip prefix-list prefix-name
   ```
2. Create a prefix list with a sequence number and a deny or permit action.
   CONFIG-NPREFIXL mode

   ```
   seq sequence-number {deny | permit} ip-prefix [ge min-prefix-length] [le
   max-prefix-length]
   ```

   The optional parameters are:

   - `ge min-prefix-length`: the minimum prefix length to match (from 0 to 32).

- `le` *`max-prefix-length`*: the maximum prefix length to match (from 0 to 32).

Example of Assigning Sequence Numbers to Filters

If you want to forward all routes that do not match the prefix list criteria, configure a prefix list filter to permit all routes (`permit 0.0.0.0/0 le 32`). The "permit all" filter must be the last filter in your prefix list. To permit the default route only, enter `permit 0.0.0.0/0`.

The following example shows how the `seq` command orders the filters according to the sequence number assigned. In the example, filter 20 was configured before filter 15 and 12, but the `show config` command displays the filters in the correct order.

```
Dell(conf-nprefixl)#seq 20 permit 0.0.0.0/0 le 32
Dell(conf-nprefixl)#seq 12 deny 134.23.0.0 /16
Dell(conf-nprefixl)#seq 15 deny 120.23.14.0 /8 le 16
Dell(conf-nprefixl)#show config
!
ip prefix-list juba
  seq 12 deny 134.23.0.0/16
  seq 15 deny 120.0.0.0/8 le 16
  seq 20 permit 0.0.0.0/0 le 32
Dell(conf-nprefixl)#
```

NOTE: The last line in the prefix list Juba contains a "permit all" statement. By including this line in a prefix list, you specify that all routes not matching any criteria in the prefix list are forwarded.

To delete a filter, use the `no seq` *`sequence-number`* command in PREFIX LIST mode.

If you are creating a standard prefix list with only one or two filters, you can let Dell Networking OS assign a sequence number based on the order in which the filters are configured. The Dell Networking OS assigns filters in multiples of five.

Creating a Prefix List Without a Sequence Number

To create a filter without a specified sequence number, use the following commands.

1. Create a prefix list and assign it a unique name.
   CONFIGURATION mode

   `ip prefix-list` *`prefix-name`*
2. Create a prefix list filter with a deny or permit action.
   CONFIG-NPREFIXL mode

   `{deny | permit}` *`ip-prefix`* `[ge` *`min-prefix-length`*`] [le` *`max-prefix-length`*`]`

   The optional parameters are:

   - `ge` *`min-prefix-length`*: is the minimum prefix length to be matched (0 to 32).
   - `le` *`max-prefix-length`*: is the maximum prefix length to be matched (0 to 32).

Example of Creating a Filter with Dell Networking OS-Assigned Sequence Numbers

The example shows a prefix list in which the sequence numbers were assigned by the software. The filters were assigned sequence numbers based on the order in which they were configured (for example, the first filter was given the lowest sequence number). The `show config` command in PREFIX LIST mode displays the two filters with the sequence numbers 5 and 10.

```
Dell(conf-nprefixl)#permit 123.23.0.0 /16
Dell(conf-nprefixl)#deny 133.24.56.0 /8
Dell(conf-nprefixl)#show conf
!
ip prefix-list awe
  seq 5 permit 123.23.0.0/16
  seq 10 deny 133.0.0.0/8
Dell(conf-nprefixl)#
```

To delete a filter, enter the `show config` command in PREFIX LIST mode and locate the sequence number of the filter you want to delete, then use the `no seq sequence-number` command in PREFIX LIST mode.

Viewing Prefix Lists

To view all configured prefix lists, use the following commands.

- Show detailed information about configured prefix lists.

  EXEC Privilege mode

  `show ip prefix-list detail [`*`prefix-name`*`]`
- Show a table of summarized information about configured Prefix lists.

  EXEC Privilege mode

  `show ip prefix-list summary [`*`prefix-name`*`]`

Examples of the `show ip prefix-list detail` and `show ip prefix-list summary` Commands

The following example shows the `show ip prefix-list detail` command.

```
Dell>show ip prefix detail
Prefix-list with the last deletion/insertion: filter_ospf
ip prefix-list filter_in:
count: 3, range entries: 3, sequences: 5 - 10
   seq 5 deny 1.102.0.0/16 le 32 (hit count: 0)
   seq 6 deny 2.1.0.0/16 ge 23 (hit count: 0)
   seq 10 permit 0.0.0.0/0 le 32 (hit count: 0)
ip prefix-list filter_ospf:
count: 4, range entries: 1, sequences: 5 - 10
   seq 5 deny 100.100.1.0/24 (hit count: 0)
   seq 6 deny 200.200.1.0/24 (hit count: 0)
   seq 7 deny 200.200.2.0/24 (hit count: 0)
   seq 10 permit 0.0.0.0/0 le 32 (hit count: 0)
```

The following example shows the `show ip prefix-list summary` command.

```
Dell>
Dell>show ip prefix summary
Prefix-list with the last deletion/insertion: filter_ospf
ip prefix-list filter_in:
count: 3, range entries: 3, sequences: 5 - 10
ip prefix-list filter_ospf:
count: 4, range entries: 1, sequences: 5 - 10
Dell>
```

Applying a Prefix List for Route Redistribution

To pass traffic through a configured prefix list, use the prefix list in a `route redistribution` command.
Apply the prefix list to all traffic redistributed into the routing process. The traffic is either forwarded or dropped, depending on the criteria and actions specified in the prefix list.

To apply a filter to routes in RIP, use the following commands.

- Enter RIP mode.

  CONFIGURATION mode

  ```
  router rip
  ```
- Apply a configured prefix list to incoming routes. You can specify an interface.

  If you enter the name of a nonexistent prefix list, all routes are forwarded.

  CONFIG-ROUTER-RIP mode

  `distribute-list` *`prefix-list-name`* `in [`*`interface`*`]`
- Apply a configured prefix list to outgoing routes. You can specify an interface or type of route.

  If you enter the name of a non-existent prefix list, all routes are forwarded.

  CONFIG-ROUTER-RIP mode

  `distribute-list` *`prefix-list-name`* `out [`*`interface`* `| connected | static | ospf]`

Example of Viewing Configured Prefix Lists (ROUTER RIP mode)

To view the configuration, use the `show config` command in ROUTER RIP mode, or the `show running-config rip` command in EXEC mode.

```
Dell(conf-router_rip)#show config
!
router rip
  distribute-list prefix juba out
  network 10.0.0.0
Dell(conf-router_rip)#router ospf 34
```

Applying a Filter to a Prefix List (OSPF)

To apply a filter to routes in open shortest path first (OSPF), use the following commands.

- Enter OSPF mode.

  CONFIGURATION mode

  ```
  router ospf
  ```
- Apply a configured prefix list to incoming routes. You can specify an interface.

  If you enter the name of a non-existent prefix list, all routes are forwarded.

  CONFIG-ROUTER-OSPF mode

  `distribute-list` *`prefix-list-name`* `in [`*`interface`*`]`
- Apply a configured prefix list to incoming routes. You can specify which type of routes are affected.

  If you enter the name of a non-existent prefix list, all routes are forwarded.

CONFIG-ROUTER-OSPF mode

`distribute-list` *`prefix-list-name`* `out [connected | rip | static]`

Example of Viewing Configured Prefix Lists (ROUTER OSPF mode)

To view the configuration, use the `show config` command in ROUTER OSPF mode, or the `show running-config ospf` command in EXEC mode.

```
Dell(conf-router_ospf)#show config
!
router ospf 34
  network 10.2.1.1 255.255.255.255 area 0.0.0.1
  distribute-list prefix awe in
Dell(conf-router_ospf)#
```

ACL Resequencing

ACL resequencing allows you to re-number the rules and remarks in an access or prefix list.

The placement of rules within the list is critical because packets are matched against rules in sequential order. To order new rules using the current numbering scheme, use resequencing whenever there is no opportunity.

For example, the following table contains some rules that are numbered in increments of 1. You cannot place new rules between these packets, so apply resequencing to create numbering space, as shown in the second table. In the same example, apply resequencing if more than two rules must be placed between rules 7 and 10.

You can resequence IPv4 and IPv6 ACLs, prefixes, and MAC ACLs. No CAM writes happen as a result of resequencing, so there is no packet loss; the behavior is similar Hot-lock ACLs.

NOTE: ACL resequencing does not affect the rules, remarks, or order in which they are applied. Resequencing merely renumbers the rules so that you can place new rules within the list as needed.

Table 5. ACL Resequencing

Rules	Resquencing
Rules Before Resequencing:	seq 5 permit any host 1.1.1.1
	seq 6 permit any host 1.1.1.2
	seq 7 permit any host 1.1.1.3
	seq 10 permit any host 1.1.1.4
Rules After Resequencing:	seq 5 permit any host 1.1.1.1
	seq 10 permit any host 1.1.1.2
	seq 15 permit any host 1.1.1.3
	seq 20 permit any host 1.1.1.4

Resequencing an ACL or Prefix List

Resequencing is available for IPv4 and IPv6 ACLs, prefix lists, and MAC ACLs.
To resequence an ACL or prefix list, use the following commands. You must specify the list name, starting number, and increment when using these commands.

- IPv4, IPv6, or MAC ACL

  EXEC mode

  ```
  resequence access-list {ipv4 | ipv6 | mac} {access-list-name StartingSeqNum
  Step-to-Increment}
  ```
- IPv4 or IPv6 prefix-list

  EXEC mode

  ```
  resequence prefix-list {ipv4 | ipv6} {prefix-list-name StartingSeqNum Step-
  to-Increment}
  ```

Examples of Resequencing ACLs When Remarks and Rules Have the Same Number or have Different Numbers

Remarks and rules that originally have the same sequence number have the same sequence number after you apply the `resequence` command.

The example shows the resequencing of an IPv4 access-list beginning with the number 2 and incrementing by 2.

```
Dell(config-ext-nacl)# show config
!
ip access-list extended test
remark 4 XYZ
remark 5 this remark corresponds to permit any host 1.1.1.1
seq 5 permit ip any host 1.1.1.1
remark 9 ABC
remark 10 this remark corresponds to permit ip any host 1.1.1.2
seq 10 permit ip any host 1.1.1.2
seq 15 permit ip any host 1.1.1.3
seq 20 permit ip any host 1.1.1.4
Dell# end
Dell# resequence access-list ipv4 test 2 2
Dell# show running-config acl
!
ip access-list extended test
remark 2 XYZ
remark 4 this remark corresponds to permit any host 1.1.1.1
seq 4 permit ip any host 1.1.1.1
remark 6 this remark has no corresponding rule
remark 8 this remark corresponds to permit ip any host 1.1.1.2
seq 8 permit ip any host 1.1.1.2
seq 10 permit ip any host 1.1.1.3
seq 12 permit ip any host 1.1.1.4
```

Remarks that do not have a corresponding rule are incremented as a rule. These two mechanisms allow remarks to retain their original position in the list. The following example shows remark 10 corresponding to rule 10 and as such, they have the same number before and after the command is entered. Remark 4 is incremented as a rule, and all rules have retained their original positions.

```
Dell(config-ext-nacl)# show config
!
ip access-list extended test
remark 4 XYZ
remark 5 this remark corresponds to permit any host 1.1.1.1
seq 5 permit ip any host 1.1.1.1
remark 9 ABC
remark 10 this remark corresponds to permit ip any host 1.1.1.2
seq 10 permit ip any host 1.1.1.2
```

```
seq 15 permit ip any host 1.1.1.3
seq 20 permit ip any host 1.1.1.4
Dell# end
Dell# resequence access-list ipv4 test 2 2
Dell# show running-config acl
!
ip access-list extended test
remark 2 XYZ
remark 4 this remark corresponds to permit any host 1.1.1.1
seq 4 permit ip any host 1.1.1.1
remark 6 this remark has no corresponding rule
remark 8 this remark corresponds to permit ip any host 1.1.1.2
seq 8 permit ip any host 1.1.1.2
seq 10 permit ip any host 1.1.1.3
seq 12 permit ip any host 1.1.1.4
```

Route Maps

Route maps are supported on S6000 platform.

Similar to ACLs and prefix lists, route maps are composed of a series of commands that contain a matching criterion and an action; however, route maps can change the packets meeting the criterion. ACLs and prefix lists can only drop or forward the packet or traffic. Route maps process routes for route redistribution. For example, a route map can be called to filter only specific routes and to add a metric.

Route maps also have an “implicit deny.” Unlike ACLs and prefix lists; however, where the packet or traffic is dropped, in route maps, if a route does not match any of the route map conditions, the route is not redistributed.

Implementation Information

The Dell Networking OS implementation of route maps allows route maps with the `no match` or `no set` commands. When there is `no match` command, all traffic matches the route map and the `set` command applies.

Flow-Based Monitoring Support for ACLs

Flow-based monitoring is supported on the S6000 platform.

Flow-based monitoring conserves bandwidth by monitoring only the specified traffic instead of all traffic on the interface. It is available for Layer 2 and Layer 3 ingress traffic. You can specify traffic using standard or extended access-lists. This mechanism copies incoming packets that matches the ACL rules applied on the ingress port and forwards (mirrors) them to another port. The source port is the monitored port (MD) and the destination port is the monitoring port (MG).

The port mirroring application maintains and performs all the monitoring operations on the chassis. ACL information is sent to the ACL manager, which in turn notifies the ACL agent to add entries in the CAM area. Duplicate entries in the ACL are not saved.

When a packet arrives at a port that is being monitored, the packet is validated against the configured ACL rules. If the packet matches an ACL rule, the system examines the corresponding flow processor to perform the action specified for that port. If the mirroring action is set in the flow processor entry, the destination port details, to which the mirrored information must be sent, are sent to the destination port.

When a stack unit is reset or a stack unit undergoes a failure, the ACL agent registers with the port mirroring application. The port mirroring utility downloads the monitoring configuration to the ACL agent. The interface manager notifies the port mirroring application about the removal of an interface when an ACL entry associated with that interface to is deleted.

Behavior of Flow-Based Monitoring

Activate flow-based monitoring for a monitoring session by entering the `flow-based enable` command in the Monitor Session mode. When you enable this capability, traffic with particular flows that are traversing through the ingress interfaces are examined, and appropriate ACLs can be applied in the ingress direction. By default, flow-based monitoring is not enabled.

You must specify the monitor option with the `permit, deny`, or `seq` command for ACLs that are assigned to the source or the monitored port (MD) to enable the evaluation and replication of traffic that is traversing to the destination port. Enter the keyword `monitor` with the `seq`, `permit`, or `deny` command for the ACL rules to allow or drop IPv4, IPv6, ARP, UDP, EtherType, ICMP, and TCP packets. The ACL rule describes the traffic that you want to monitor, and the ACL in which you are creating the rule will be applied to the monitored interface. Flow monitoring is supported for standard and extended IPv4 ACLs, standard and extended IPv6 ACLs, and standard and extended MAC ACLs.

CONFIG-STD-NACL mode

```
seq sequence-number {deny | permit} {source [mask] | any | host ip-address}
[count [byte]] [order] [fragments] [log [threshold-in-msgs count]] [monitor]
```

If the number of monitoring sessions increases, inter-process communication (IPC) bandwidth utilization will be high. The ACL manager might require a large bandwidth when you assign an ACL, with many entries, to an interface.

The ACL agent module saves monitoring details in its local database and also in the CAM region to monitor packets that match the specified criterion. The ACL agent maintains data on the source port, the destination port, and the endpoint to which the packet must be forwarded when a match occurs with the ACL entry.

If you configure the `flow-based enable` command and do not apply an ACL on the source port or the monitored port, both flow-based monitoring and port mirroring do not function. Flow-based monitoring is supported only for ingress traffic and not for egress packets.

The port mirroring application maintains a database that contains all monitoring sessions (including port monitor sessions). It has information regarding the sessions that are enabled for flow-based monitoring and those sessions that are not enabled for flow-based monitoring. It downloads monitoring configuration to the ACL agent whenever the ACL agent is registered with the port mirroring application or when flow-based monitoring is enabled.

The `show monitor session` *`session-id`* command has been enhanced to display the Type field in the output, which indicates whether a particular session is enabled for flow-monitoring.

Example Output of the `show` Command

```
Dell(conf-mon-sess-0)#do show monitor session 0

SessID  Source         Destination         Dir  Mode  Source IP      Dest IP
------  ------         -----------         ---  ----  ---------      --------
```

```
  0     Te 0/0            Te 0/2           rx    Flow     N/A             N/
A
```

The `show config` command has been modified to display monitoring configuration in a particular session.

Example Output of the `show` Command

```
(conf-mon-sess-11)#show config
    !
    monitor session 11
                          flow-based enable
      source GigabitEthernet 13/0 destination GigabitEthernet 13/1 direction
both
```

The `show ip | mac | ipv6 accounting` commands have been enhanced to display whether monitoring is enabled for traffic that matches with the rules of the specific ACL.

Example Output of the `show` Command

```
Dell# show ip accounting access-list
!
Extended Ingress IP access list kar on GigabitEthernet 10/0
Total cam count 1
   seq 5 permit ip 192.168.20.0/24 173.168.20.0/24 monitor

Dell#show ipv6 accounting access-list
    !
    Ingress IPv6 access list kar on GigabitEthernet 10/0
    Total cam count 1
     seq 5 permit ipv6 22::/24 33::/24 monitor
```

Enabling Flow-Based Monitoring

Flow-based monitoring is supported on the S6000 platform.
Flow-based monitoring conserves bandwidth by monitoring only specified traffic instead of all traffic on the interface. This feature is particularly useful when looking for malicious traffic. It is available for Layer 2 and Layer 3 ingress and egress traffic. You can specify traffic using standard or extended access-lists.

1. Enable flow-based monitoring for a monitoring session.

   MONITOR SESSION mode

   ```
   flow-based enable
   ```

2. Define access-list rules that include the keyword `monitor`. Dell Networking OS only considers port monitoring traffic that matches rules with the keyword `monitor`.

   CONFIGURATION mode

   ```
   ip access-list
   ```

   For more information, see Access Control Lists (ACLs).

3. Apply the ACL to the monitored port.

   INTERFACE mode

   ```
   ip access-group access-list
   ```

Example of the `flow-based enable` Command

To view an access-list that you applied to an interface, use the `show ip accounting access-list` command from EXEC Privilege mode.

```
Dell(conf)#monitor session 0
Dell(conf-mon-sess-0)#flow-based enable
Dell(conf)#ip access-list ext testflow
Dell(config-ext-nacl)#seq 5 permit icmp any any count bytes monitor
Dell(config-ext-nacl)#seq 10 permit ip 102.1.1.0/24 any count bytes monitor
Dell(config-ext-nacl)#seq 15 deny udp any any count bytes
Dell(config-ext-nacl)#seq 20 deny tcp any any count bytes
Dell(config-ext-nacl)#exit
Dell(conf)#interface gig 1/1
Dell(conf-if-gi-1/1)#ip access-group testflow in
Dell(conf-if-gi-1/1)#show config
!
interface GigabitEthernet 1/1
  ip address 10.11.1.254/24
  ip access-group testflow in
  shutdown
Dell(conf-if-gi-1/1)#exit
Dell(conf)#do show ip accounting access-list testflow
!
Extended Ingress IP access list testflow on GigabitEthernet 1/1
Total cam count 4
  seq 5 permit icmp any any monitor count bytes (0 packets 0 bytes)
  seq 10 permit ip 102.1.1.0/24 any monitor count bytes (0 packets 0 bytes)
  seq 15 deny udp any any count bytes (0 packets 0 bytes)
  seq 20 deny tcp any any count bytes (0 packets 0 bytes)
Dell(conf)#do show monitor session 0
ct-maa-s4820-2(conf-mon-sess-0)#do show monitor session 0

SessID  Source         Destination         Dir  Mode  Source IP      Dest IP
------  ------         -----------         ---  ----  ---------      --------
  0     Te 0/0            Te 0/2            rx   Flow     N/A            N/
A
```

7

Bidirectional Forwarding Detection (BFD)

Bidirectional forwarding detection (BFD) is supported only on the S6000 platform.

BFD is a protocol that is used to rapidly detect communication failures between two adjacent systems. It is a simple and lightweight replacement for existing routing protocol link state detection mechanisms. It also provides a failure detection solution for links on which no routing protocol is used.

BFD is a simple hello mechanism. Two neighboring systems running BFD establish a session using a three-way handshake. After the session has been established, the systems exchange periodic control packets at sub-second intervals. If a system does not receive a hello packet within a specified amount of time, routing protocols are notified that the forwarding path is down.

BFD provides forwarding path failure detection times on the order of milliseconds rather than seconds as with conventional routing protocol hellos. It is independent of routing protocols, and as such, provides a consistent method of failure detection when used across a network. Networks converge faster because BFD triggers link state changes in the routing protocol sooner and more consistently because BFD eliminates the use of multiple protocol-dependent timers and methods.

BFD also carries less overhead than routing protocol hello mechanisms. Control packets can be encapsulated in any form that is convenient, and, on Dell Networking routers, BFD agents maintain sessions that reside on the line card, which frees resources on the route processor module (RPM). Only session state changes are reported to the BFD Manager (on the RPM), which in turn notifies the routing protocols that are registered with it.

BFD is an independent and generic protocol, which all media, topologies, and routing protocols can support using any encapsulation. Dell Networking has implemented BFD at Layer 3 and with user datagram protocol (UDP) encapsulation. BFD functionality will be implemented in phases. On the S6000 platform, BFD is supported on dynamic routing protocols such as VRRP, OSPF, OSPFv3, IS-IS, and BGP.

How BFD Works

Two neighboring systems running BFD establish a session using a three-way handshake.

After the session has been established, the systems exchange control packets at agreed upon intervals. In addition, systems send a control packet anytime there is a state change or change in a session parameter. These control packets are sent without regard to transmit and receive intervals.

NOTE: The Dell Networking Operating System (OS) does not support multi-hop BFD sessions.

If a system does not receive a control packet within an agreed-upon amount of time, the BFD agent changes the session state to Down. It then notifies the BFD manager of the change and sends a control packet to the neighbor that indicates the state change (though it might not be received if the link or receiving interface is faulty). The BFD manager notifies the routing protocols that are registered with it (clients) that the forwarding path is down and a link state change is triggered in all protocols.

NOTE: A session state change from Up to Down is the only state change that triggers a link state change in the routing protocol client.

BFD Packet Format

Control packets are encapsulated in user datagram protocol (UDP) packets. The following illustration shows the complete encapsulation of a BFD control packet inside an IPv4 packet.

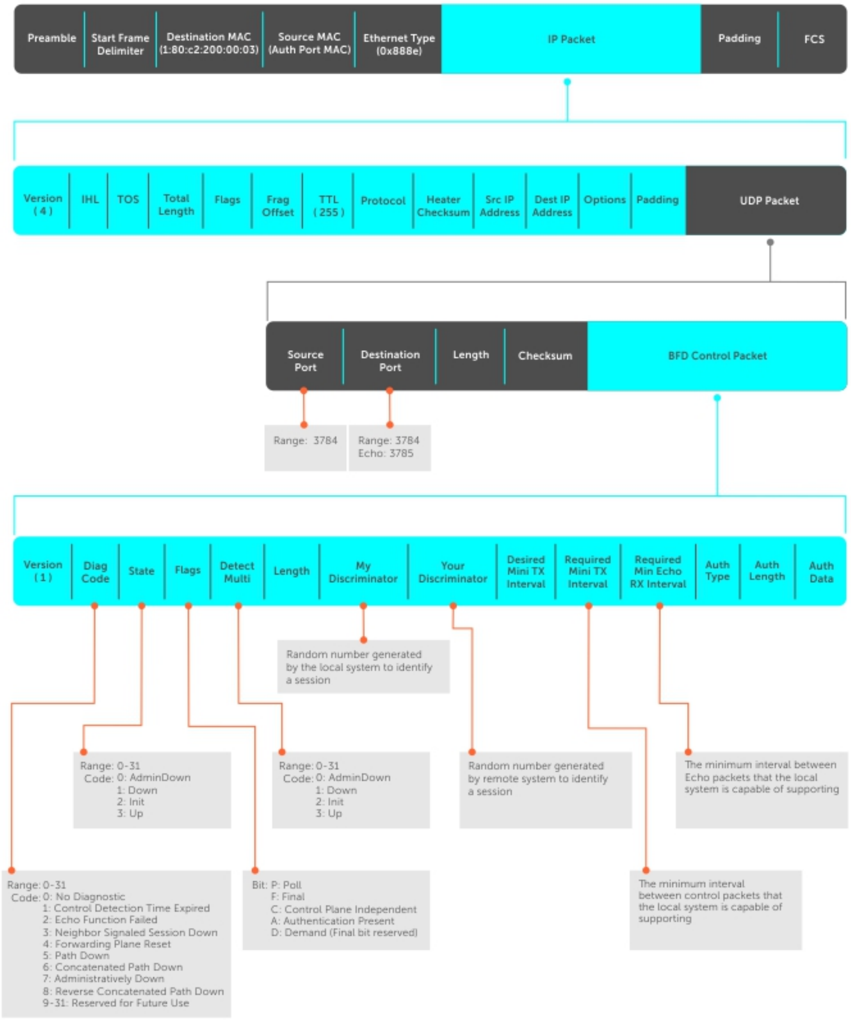

Preamble
Start Frame Delimiter
Destination MAC (1:80:c2:200:00:03)
Source MAC (Auth Port MAC)
Ethernet Type (0x888e)
IP Packet
Padding
FCS
Version (4)
IHL
TOS
Total Length
Flags
Frag Offset
TTL (255)
Protocol
Heater Checksum
Src IP Address
Dest IP Address
Options
Padding
UDP Packet
Source Port
Destination Port
Length
Checksum
BFD Control Packet
Range: 3784
Range: 3784 Echo: 3785
Version (1)
Diag Code
State
Flags
Detect Multi
Length
My Discriminator
Your Discriminator
Desired Mini TX Interval
Required Mini TX Interval
Required Min Echo RX Interval
Auth Type
Auth Length
Auth Data
Random number generated by the local system to identify a session
Range: 0-31 Code: 0: AdminDown 1: Down 2: Init 3: Up
Range: 0-31 Code: 0: AdminDown 1: Down 2: Init 3: Up
Random number generated by remote system to identify a session
The minimum interval between Echo packets that the local system is capable of supporting
Range: 0-31 Code: 0: No Diagnostic 1: Control Detection Time Expired 2: Echo Function Failed 3: Neighbor Signaled Session Down 4: Forwarding Plane Reset 5: Path Down 6: Concatenated Path Down 7: Administratively Down 8: Reverse Concatenated Path Down 9-31: Reserved for Future Use
Bit: P: Poll F: Final C: Control Plane Independent A: Authentication Present D: Demand (Final bit reserved)
The minimum interval between control packets that the local system is capable of supporting

Figure 8. BFD in IPv4 Packet Format

Field	Description
Diagnostic Code	The reason that the last session failed.
State	The current local session state. Refer to BFD Sessions.
Flag	A bit that indicates packet function. If the poll bit is set, the receiving system must respond as soon as possible, without regard to its transmit interval. The responding

Field	Description
	system clears the poll bit and sets the final bit in its response. The poll and final bits are used during the handshake and in Demand mode (refer to BFD Sessions). **NOTE:** Dell Networking OS does not currently support multi-point sessions, Demand mode, authentication, or control plane independence; these bits are always clear.
Detection Multiplier	The number of packets that must be missed in order to declare a session down.
Length	The entire length of the BFD packet.
My Discriminator	A random number generated by the local system to identify the session.
Your Discriminator	A random number generated by the remote system to identify the session. Discriminator values are necessary to identify the session to which a control packet belongs because there can be many sessions running on a single interface.
Desired Min TX Interval	The minimum rate at which the local system would like to send control packets to the remote system.
Required Min RX Interval	The minimum rate at which the local system would like to receive control packets from the remote system.
Required Min Echo RX	The minimum rate at which the local system would like to receive echo packets. **NOTE:** Dell Networking OS does not currently support the echo function.
Authentication Type, Authentication Length, Authentication Data	An optional method for authenticating control packets. **NOTE:** Dell Networking OS does not currently support the BFD authentication function.

Two important parameters are calculated using the values contained in the control packet.

Transmit interval	Transmit interval is the agreed-upon rate at which a system sends control packets. Each system has its own transmit interval, which is the greater of the last received remote Desired TX Interval and the local Required Min RX Interval.
Detection time	Detection time is the amount of time that a system does not receive a control packet, after which the system determines that the session has failed. Each system has its own detection time. • In Asynchronous mode: Detection time is the remote Detection Multiplier multiplied by greater of the remote Desired TX Interval and the local Required Min RX Interval. • In Demand mode: Detection time is the local Detection Multiplier multiplied by the greater of the local Desired Min TX and the remote Required Min RX Interval.

BFD Sessions

BFD must be enabled on both sides of a link in order to establish a session.

The two participating systems can assume either of two roles:

Active	The active system initiates the BFD session. Both systems can be active for the same session.
Passive	The passive system does not initiate a session. It only responds to a request for session initialization from the active system.

A BFD session has two modes:

Asynchronous mode	In Asynchronous mode, both systems send periodic control messages at an agreed upon interval to indicate that their session status is Up.'
Demand mode	If one system requests Demand mode, the other system stops sending periodic control packets; it only sends a response to status inquiries from the Demand mode initiator. Either system (but not both) can request Demand mode at any time.

NOTE: Dell Networking OS supports Asynchronous mode only.

A session can have four states: Administratively Down, Down, Init, and Up.

Administratively Down	The local system does not participate in a particular session.
Down	The remote system is not sending control packets or at least not within the detection time for a particular session.
Init	The local system is communicating.
Up	Both systems are exchanging control packets.

The session is declared down if:

- A control packet is not received within the detection time.
- Sufficient echo packets are lost.
- Demand mode is active and a control packet is not received in response to a poll packet.

BFD Three-Way Handshake

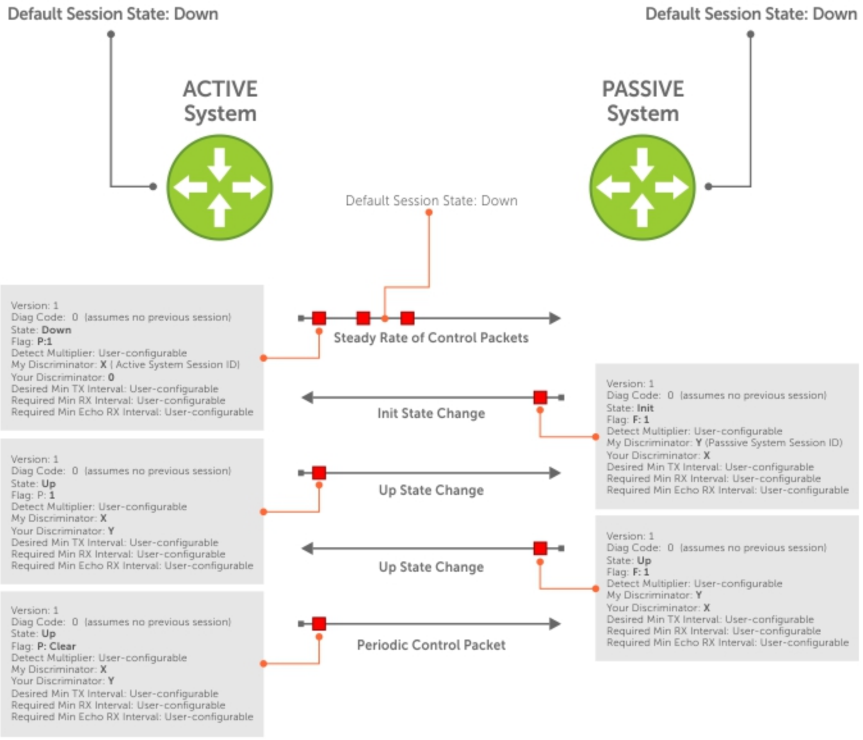
A three-way handshake must take place between the systems that participate in the BFD session.

The handshake shown in the following illustration assumes that there is one active and one passive system, and that this session is the first session established on this link. The default session state on both ports is Down.

1. The active system sends a steady stream of control packets that indicates that its session state is Down, until the passive system responds. These packets are sent at the desired transmit interval of the Active system. The Your Discriminator field is set to zero.
2. When the passive system receives any of these control packets, it changes its session state to Init and sends a response that indicates its state change. The response includes its session ID in the My Discriminator field and the session ID of the remote system in the Your Discriminator field.
3. The active system receives the response from the passive system and changes its session state to Up. It then sends a control packet indicating this state change. This is the third and final part of the

handshake. Now the discriminator values have been exchanged and the transmit intervals have been negotiated.

4. The passive system receives the control packet and changes its state to Up. Both systems agree that a session has been established. However, because both members must send a control packet — that requires a response — anytime there is a state change or change in a session parameter, the passive system sends a final response indicating the state change. After this, periodic control packets are exchanged.


Default Session State: Down
Default Session State: Down
ACTIVE System
PASSIVE System
Default Session State: Down
Steady Rate of Control Packets
Init State Change
Up State Change
Up State Change
Periodic Control Packet

Version: 1
Diag Code: 0 (assumes no previous session)
State: **Down**
Flag: **P:1**
Detect Multiplier: User-configurable
My Discriminator: **X** (Active System Session ID)
Your Discriminator: **0**
Desired Min TX Interval: User-configurable
Required Min RX Interval: User-configurable
Required Min Echo RX Interval: User-configurable

Version: 1
Diag Code: 0 (assumes no previous session)
State: **Init**
Flag: **F: 1**
Detect Multiplier: User-configurable
My Discriminator: **Y** (Passsive System Session ID)
Your Discriminator: **X**
Desired Min TX Interval: User-configurable
Required Min RX Interval: User-configurable
Required Min Echo RX Interval: User-configurable

Version: 1
Diag Code: 0 (assumes no previous session)
State: **Up**
Flag: P: **1**
Detect Multiplier: User-configurable
My Discriminator: **X**
Your Discriminator: **Y**
Desired Min TX Interval: User-configurable
Required Min RX Interval: User-configurable
Required Min Echo RX Interval: User-configurable

Version: 1
Diag Code: 0 (assumes no previous session)
State: **Up**
Flag: **F: 1**
Detect Multiplier: User-configurable
My Discriminator: **Y**
Your Discriminator: **X**
Desired Min TX Interval: User-configurable
Required Min RX Interval: User-configurable
Required Min Echo RX Interval: User-configurable

Version: 1
Diag Code: 0 (assumes no previous session)
State: **Up**
Flag: **P: Clear**
Detect Multiplier: User-configurable
My Discriminator: **X**
Your Discriminator: **Y**
Desired Min TX Interval: User-configurable
Required Min RX Interval: User-configurable
Required Min Echo RX Interval: User-configurable

Figure 9. BFD Three-Way Handshake State Changes

Session State Changes

The following illustration shows how the session state on a system changes based on the status notification it receives from the remote system. For example, if a session on a system is down and it receives a Down status notification from the remote system, the session state on the local system changes to Init.

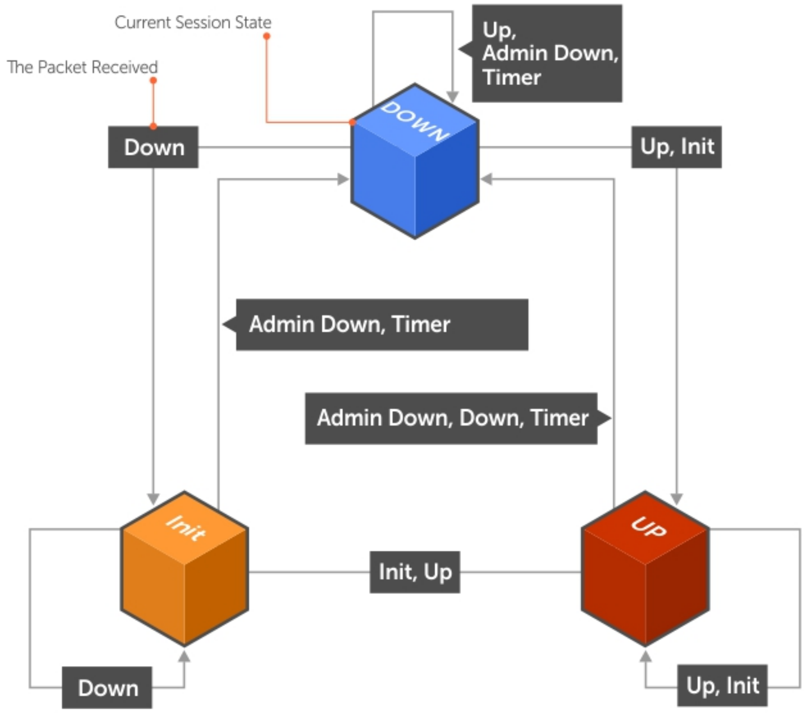

Current Session State
The Packet Received
Up,
Admin Down,
Timer
DOWN
Down
Up, Init
Admin Down, Timer
Admin Down, Down, Timer
Init
Init, Up
Up
Down
Up, Init

Figure 10. Session State Changes

Important Points to Remember

- On the platform, Dell Networking OS supports 128 sessions per stack unit at 200 minimum transmit and receive intervals with a multiplier of 3, and 64 sessions at 100 minimum transmit and receive intervals with a multiplier of 4.
- Enable BFD on both ends of a link.
- Demand mode, authentication, and the Echo function are not supported.
- BFD is not supported on multi-hop and virtual links.
- Protocol Liveness is supported for routing protocols only.

Configure BFD

This section contains the following procedures.

- Configuring BFD for Physical Ports
- Configure BFD for Static Routes
- Configure BFD for OSPF
- Configure BFD for OSPFv3

- Configure BFD for IS-IS
- Configure BFD for BGP
- Configure BFD for VRRP
- Configuring Protocol Liveness
- Troubleshooting BFD

Configure BFD for Physical Ports

Configuring BFD for physical ports is supported on the C-Series and E-Series platforms only.

BFD on physical ports is useful when you do not enable the routing protocol. Without BFD, if the remote system fails, the local system does not remove the connected route until the first failed attempt to send a packet. When you enable BFD, the local system removes the route as soon as it stops receiving periodic control packets from the remote system.

Configuring BFD for a physical port is a two-step process:

1. Enable BFD globally.
2. Establish a session with a next-hop neighbor.

Related Configuration Tasks

- Viewing Physical Port Session Parameters.
- Disabling and Re-Enabling BFD.

Enabling BFD Globally

You must enable BFD globally on both routers.

For more information about enabling BFD globally, refer to Establishing a Session on Physical Ports.

To enable the BFD globally, use the following command.

- Enable BFD globally.

  CONFIGURATION mode

  ```
  bfd enable
  ```

Example of Verifying BFD is Enabled

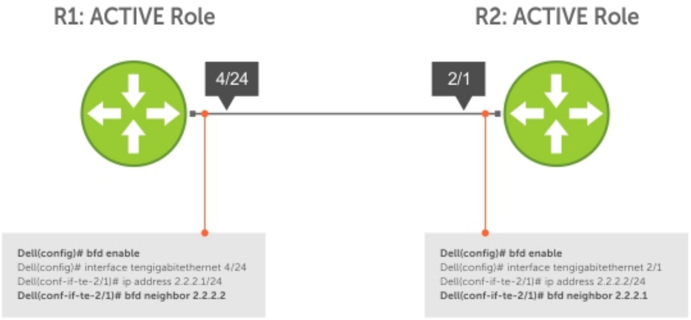
To verify that BFD is enabled globally, use the `show running bfd` command.

The bold line shows that BFD is enabled.

```
R1(conf)#bfd ?
enable                  Enable BFD protocol
protocol-liveness       Enable BFD protocol-liveness
R1(conf)#bfd enable

R1(conf)#do show running-config bfd
!
bfd enable
R1(conf)#
```

Establishing a Session on Physical Ports

To establish a session, enable BFD at the interface level on both ends of the link, as shown in the following illustration. The configuration parameters do not need to match.



```
Dell(config)# bfd enable
Dell(config)# interface tengigabitethernet 4/24
Dell(conf-if-te-2/1)# ip address 2.2.2.1/24
Dell(conf-if-te-2/1)# bfd neighbor 2.2.2.2
```

```
Dell(config)# bfd enable
Dell(config)# interface tengigabitethernet 2/1
Dell(conf-if-te-2/1)# ip address 2.2.2.2/24
Dell(conf-if-te-2/1)# bfd neighbor 2.2.2.1
```

Figure 11. Establishing a BFD Session on Physical Ports

1. Enter interface mode.

   CONFIGURATION mode

   ```
   interface
   ```
2. Assign an IP address to the interface if one is not already assigned.

   INTERFACE mode

   ```
   ip address ip-address
   ```
3. Identify the neighbor that the interface participates with the BFD session.

   INTERFACE mode

   ```
   bfd neighbor ip-address
   ```

Examples of the `show bfd neighbors` command.

To verify that the session is established, use the `show bfd neighbors` command.

The bold line shows the BFD session.

```
R1(conf-if-gi-4/24)#do show bfd neighbors
* - Active session role
Ad Dn - Admin Down
C - CLI
I - ISIS
O - OSPF
R - Static Route (RTM)
LocalAddr   RemoteAddr   Interface  State  Rx-int  Tx-int  Mult  Clients
* 2.2.2.1   2.2.2.2      Gi 4/24    Up     100     100     3     C
```

To view specific information about BFD sessions, use the `show bfd neighbors detail` command.

```
R1(conf-if-gi-4/24)#do show bfd neighbors detail

Session Discriminator: 1
Neighbor Discriminator: 1
Local Addr: 2.2.2.1
Local MAC Addr: 00:01:e8:09:c3:e5
```

```
Remote Addr: 2.2.2.2
Remote MAC Addr: 00:01:e8:06:95:a2
Int: GigabitEthernet 4/24
State: Up
Configured parameters:
  TX: 100ms, RX: 100ms, Multiplier: 3
Neighbor parameters:
  TX: 100ms, RX: 100ms, Multiplier: 3
Actual parameters:
  TX: 100ms, RX: 100ms, Multiplier: 3
Role: Active
Delete session on Down: False
Client Registered: CLI
Uptime: 00:03:57
Statistics:
  Number of packets received from neighbor: 1775
  Number of packets sent to neighbor: 1775
  Number of state changes: 1
  Number of messages from IFA about port state change: 0
  Number of messages communicated b/w Manager and Agent: 4
```

Log messages display when you configure both interfaces for BFD.

```
     R1(conf-if-gi-4/24)#00:36:01: %RPM0-P:RP2 %BFDMGR-1-BFD_STATE_CHANGE: 
Changed session state to 
Down for neighbor 2.2.2.2 on interface Gi 4/24 (diag: 0)
     00:36:02: %RPM0-P:RP2 %BFDMGR-1-BFD_STATE_CHANGE: Changed session state to 
Up for neighbor 
2.2.2.2 on interface Gi 4/24 (diag: 0)
```

Viewing Physical Port Session Parameters

BFD sessions are configured with default intervals and a default role (active). Dell Networking recommends maintaining the default values.

To view session parameters, use the `show bfd neighbors detail` command.

Example of Viewing Session Parameters

```
R1(conf-if-gi-4/24)#bfd interval 100 min_rx 100 multiplier 4 role passive
R1(conf-if-gi-4/24)#do show bfd neighbors detail

Session Discriminator: 1
Neighbor Discriminator: 1
Local Addr: 2.2.2.1
Local MAC Addr: 00:01:e8:09:c3:e5
Remote Addr: 2.2.2.2
Remote MAC Addr: 00:01:e8:06:95:a2
Int: GigabitEthernet 4/24
State: Up
Configured parameters:
  TX: 100ms, RX: 100ms, Multiplier: 4
Neighbor parameters:
  TX: 100ms, RX: 100ms, Multiplier: 3
Actual parameters:
  TX: 100ms, RX: 100ms, Multiplier: 4
Role: Passive
Delete session on Down: False
Client Registered: CLI
Uptime: 00:09:06
Statistics:
  Number of packets received from neighbor: 4092
  Number of packets sent to neighbor: 4093
  Number of state changes: 1
  Number of messages from IFA about port state change: 0
  Number of messages communicated b/w Manager and Agent: 7
```

Disabling and Re-Enabling BFD

BFD is enabled on all interfaces by default, though sessions are not created unless explicitly configured. If you disable BFD, all of the sessions on that interface are placed in an Administratively Down state (the first message example), and the remote systems are notified of the session state change (the second message example).

To disable and re-enable BFD on an interface, use the following commands.

- Disable BFD on an interface.

  INTERFACE mode

  ```
  no bfd enable
  ```
- Enable BFD on an interface.

  INTERFACE mode

  ```
  bfd enable
  ```

If you disable BFD on a local interface, this message displays:

```
     R1(conf-if-gi-4/24)#01:00:52: %RPM0-P:RP2 %BFDMGR-1-BFD_STATE_CHANGE:
Changed session state to Ad
Dn for neighbor 2.2.2.2 on interface Gi 4/24 (diag: 0)
```

If the remote system state changes due to the local state administration being down, this message displays:

```
     R2>01:32:53: %RPM0-P:RP2 %BFDMGR-1-BFD_STATE_CHANGE: Changed session state
to Down for neighbor
2.2.2.1 on interface Gi 2/1 (diag: 7)
```

Configure BFD for Static Routes

Configuring BFD for static routes is supported on , , , andS6000.

BFD offers systems a link state detection mechanism for static routes. With BFD, systems are notified to remove static routes from the routing table as soon as the link state change occurs, rather than waiting until packets fail to reach their next hop.

Configuring BFD for static routes is a three-step process:

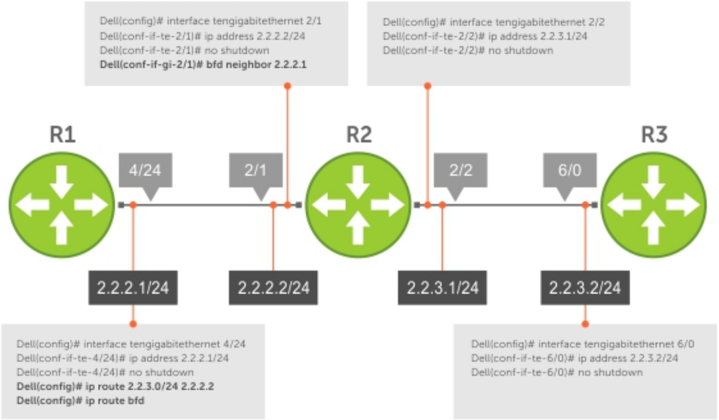
1. Enable BFD globally.
2. Configure static routes on both routers on the system (either local or remote).
3. Configure an IP route to connect BFD on the static routes using the `ip route bfd` command.

Related Configuration Tasks

- Changing Static Route Session Parameters
- Disabling BFD for Static Routes

Establishing Sessions for Static Routes

Sessions are established for all neighbors that are the next hop of a static route.



```
Dell(config)# interface tengigabitethernet 2/1
Dell(conf-if-te-2/1)# ip address 2.2.2.2/24
Dell(conf-if-te-2/1)# no shutdown
Dell(conf-if-gi-2/1)# bfd neighbor 2.2.2.1
```

```
Dell(config)# interface tengigabitethernet 2/2
Dell(conf-if-te-2/2)# ip address 2.2.3.1/24
Dell(conf-if-te-2/2)# no shutdown
```

```
Dell(config)# interface tengigabitethernet 4/24
Dell(conf-if-te-4/24)# ip address 2.2.2.1/24
Dell(conf-if-te-4/24)# no shutdown
Dell(config)# ip route 2.2.3.0/24 2.2.2.2
Dell(config)# ip route bfd
```

```
Dell(config)# interface tengigabitethernet 6/0
Dell(conf-if-te-6/0)# ip address 2.2.3.2/24
Dell(conf-if-te-6/0)# no shutdown
```

Figure 12. Establishing Sessions for Static Routes

To establish a BFD session, use the following command.

- Establish BFD sessions for all neighbors that are the next hop of a static route.
  CONFIGURATION mode

  ```
  ip route bfd
  ```

Example of the `show bfd neighbors` Command to Verify Static Routes

To verify that sessions have been created for static routes, use the `show bfd neighbors` command.

```
R1(conf)#ip route 2.2.3.0/24 2.2.2.2
R1(conf)#ip route bfd
R1(conf)#do show bfd neighbors

* - Active session role
Ad Dn - Admin Down
C - CLI
I - ISIS
O - OSPF
R - Static Route (RTM)
LocalAddr RemoteAddr Interface State Rx-int Tx-int Mult Clients
2.2.2.1   2.2.2.2    Gi 4/24   Up    100    100    4    R
```

To view detailed session information, use the `show bfd neighbors detail` command, as shown in the examples in Displaying BFD for BGP Information.

Changing Static Route Session Parameters

BFD sessions are configured with default intervals and a default role.
The parameters you can configure are: Desired TX Interval, Required Min RX Interval, Detection Multiplier, and system role. These parameters are configured for all static routes. If you change a parameter, the change affects all sessions for static routes.

To change parameters for static route sessions, use the following command .

- Change parameters for all static route sessions.

  CONFIGURATION mode

  ```
  ip route bfd interval milliseconds min_rx milliseconds multiplier value role
  [active | passive]
  ```

To view session parameters, use the `show bfd neighbors detail` command, as shown in the examples in Displaying BFD for BGP Information.

Disabling BFD for Static Routes

If you disable BFD, all static route BFD sessions are torn down.
A final Admin Down packet is sent to all neighbors on the remote systems, and those neighbors change to the Down state.

To disable BFD for static routes, use the following command.

- Disable BFD for static routes.

  CONFIGURATION mode

  ```
  no ip route bfd
  ```

Configure BFD for OSPF

BFD for OSPF is only supported on the S6000 platform.

When using BFD with OSPF, the OSPF protocol registers with the BFD manager on the RPM. BFD sessions are established with all neighboring interfaces participating in OSPF. If a neighboring interface fails, the BFD agent on the line card notifies the BFD manager, which in turn notifies the OSPF protocol that a link state change occurred.

Configuring BFD for OSPF is a two-step process:

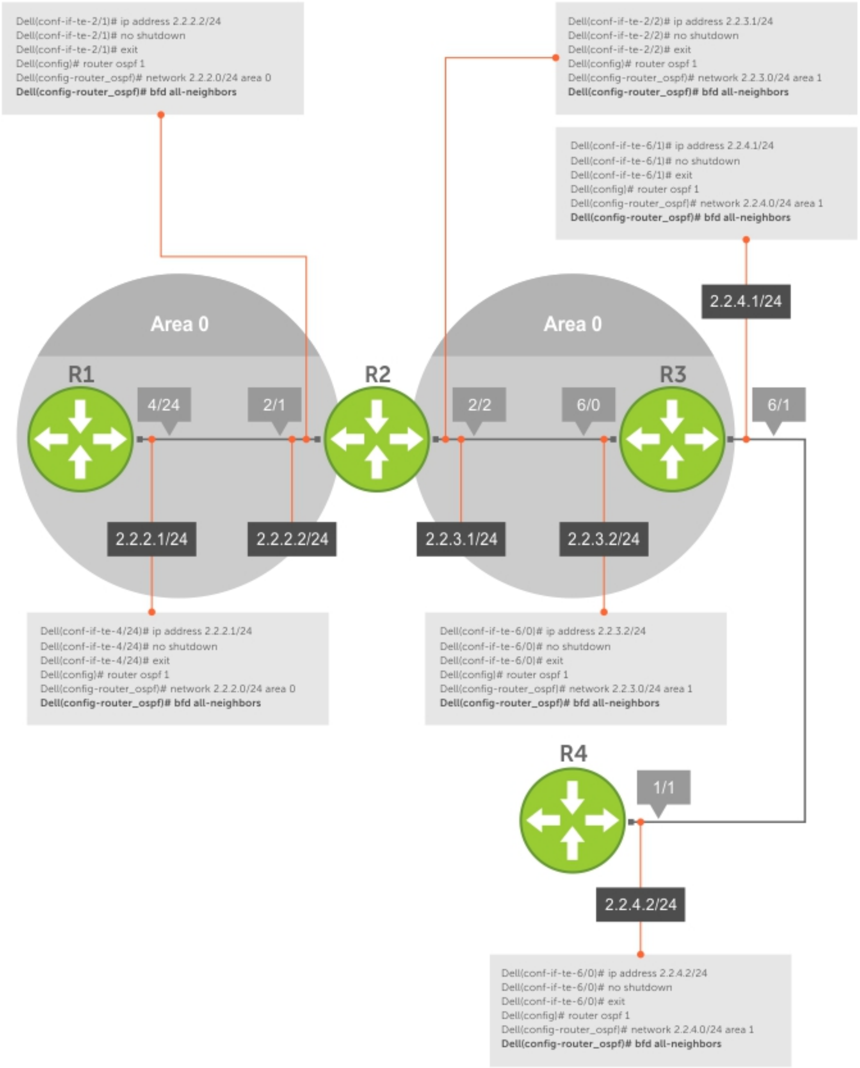
1. Enable BFD globally.
2. Establish sessions with OSPF neighbors.

Related Configuration Tasks

- Changing OSPF Session Parameters
- Disabling BFD for OSPF

Establishing Sessions with OSPF Neighbors

BFD sessions can be established with all OSPF neighbors at once or sessions can be established with all neighbors out of a specific interface. Sessions are only established when the OSPF adjacency is in the Full state.



```
Dell(conf-if-te-2/1)# ip address 2.2.2.2/24
Dell(conf-if-te-2/1)# no shutdown
Dell(conf-if-te-2/1)# exit
Dell(config)# router ospf 1
Dell(config-router_ospf)# network 2.2.2.0/24 area 0
Dell(config-router_ospf)# bfd all-neighbors
```

```
Dell(conf-if-te-2/2)# ip address 2.2.3.1/24
Dell(conf-if-te-2/2)# no shutdown
Dell(conf-if-te-2/2)# exit
Dell(config)# router ospf 1
Dell(config-router_ospf)# network 2.2.3.0/24 area 1
Dell(config-router_ospf)# bfd all-neighbors
```

```
Dell(conf-if-te-6/1)# ip address 2.2.4.1/24
Dell(conf-if-te-6/1)# no shutdown
Dell(conf-if-te-6/1)# exit
Dell(config)# router ospf 1
Dell(config-router_ospf)# network 2.2.4.0/24 area 1
Dell(config-router_ospf)# bfd all-neighbors
```

```
Dell(conf-if-te-4/24)# ip address 2.2.2.1/24
Dell(conf-if-te-4/24)# no shutdown
Dell(conf-if-te-4/24)# exit
Dell(config)# router ospf 1
Dell(config-router_ospf)# network 2.2.2.0/24 area 0
Dell(config-router_ospf)# bfd all-neighbors
```

```
Dell(conf-if-te-6/0)# ip address 2.2.3.2/24
Dell(conf-if-te-6/0)# no shutdown
Dell(conf-if-te-6/0)# exit
Dell(config)# router ospf 1
Dell(config-router_ospf)# network 2.2.3.0/24 area 1
Dell(config-router_ospf)# bfd all-neighbors
```

```
Dell(conf-if-te-6/0)# ip address 2.2.4.2/24
Dell(conf-if-te-6/0)# no shutdown
Dell(conf-if-te-6/0)# exit
Dell(config)# router ospf 1
Dell(config-router_ospf)# network 2.2.4.0/24 area 1
Dell(config-router_ospf)# bfd all-neighbors
```

Figure 13. Establishing Sessions with OSPF Neighbors

To establish BFD with all OSPF neighbors or with OSPF neighbors on a single interface, use the following commands.

- Establish sessions with all OSPF neighbors.

  ROUTER-OSPF mode

  ```
  bfd all-neighbors
  ```
- Establish sessions with OSPF neighbors on a single interface.

INTERFACE mode

```
ip ospf bfd all-neighbors
```

Example of Verifying Sessions with OSPF Neighbors

To view the established sessions, use the `show bfd neighbors` command.

The bold line shows the OSPF BFD sessions.

```
R2(conf-router_ospf)#bfd all-neighbors
R2(conf-router_ospf)#do show bfd neighbors

*      - Active session role
Ad Dn - Admin Down
C      - CLI
I      - ISIS
O      - OSPF
R      - Static Route (RTM)

LocalAddr  RemoteAddr Interface State Rx-int Tx-int Mult Clients
* 2.2.2.2  2.2.2.1    Gi 2/1    Up    100    100    3    O
* 2.2.3.1 2.2.3.2  Gi 2/2  Up  100  100  3  O
```

Changing OSPF Session Parameters

Configure BFD sessions with default intervals and a default role.
The parameters that you can configure are: `desired tx interval`, `required min rx interval`, `detection multiplier`, and `system role`. Configure these parameters for all OSPF sessions or all OSPF sessions on a particular interface. If you change a parameter globally, the change affects all OSPF neighbors sessions. If you change a parameter at the interface level, the change affects all OSPF sessions on that interface.

To change parameters for all OSPF sessions or for OSPF sessions on a single interface, use the following commands.

- Change parameters for OSPF sessions.

  ROUTER-OSPF mode

  ```
  bfd all-neighbors interval milliseconds min_rx milliseconds multiplier value
  role [active | passive]
  ```
- Change parameters for all OSPF sessions on an interface.

  INTERFACE mode

  ```
  ip ospf bfd all-neighbors interval milliseconds min_rx milliseconds
  multiplier value role [active | passive]
  ```

To view session parameters, use the `show bfd neighbors detail` command.

Disabling BFD for OSPF

If you disable BFD globally, all sessions are torn down and sessions on the remote system are placed in a Down state.
If you disable BFD on an interface, sessions on the interface are torn down and sessions on the remote system are placed in a Down state. Disabling BFD does not trigger a change in BFD clients; a final Admin Down packet is sent before the session is terminated.

To disable BFD sessions, use the following commands.

- Disable BFD sessions with all OSPF neighbors.

  ROUTER-OSPF mode

  ```
  no bfd all-neighbors
  ```
- Disable BFD sessions with all OSPF neighbors on an interface.

  INTERFACE mode

  ```
  ip ospf bfd all-neighbors disable
  ```

Configure BFD for OSPFv3

BFD for OSPFv3 is only supported on the platform.

BFD for OSPFv3 provides support for IPV6.

Configuring BFD for OSPFv3 is a two-step process:

1. Enable BFD globally.
2. Establish sessions with OSPFv3 neighbors.

Related Configuration Tasks

- Changing OSPFv3 Session Parameters
- Disabling BFD for OSPFv3

Establishing Sessions with OSPFv3 Neighbors

You can establish BFD sessions with all OSPFv3 neighbors at once or with all neighbors out of a specific interface. Sessions are only established when the OSPFv3 adjacency is in the Full state.

To establish BFD with all OSPFv3 neighbors or with OSPFv3 neighbors on a single interface, use the following commands.

- Establish sessions with all OSPFv3 neighbors.

  ROUTER-OSPFv3 mode

  ```
  bfd all-neighbors
  ```
- Establish sessions with OSPFv3 neighbors on a single interface.

  INTERFACE mode

  ```
  ipv6 ospf bfd all-neighbors
  ```

To view the established sessions, use the `show bfd neighbors` command.

Changing OSPFv3 Session Parameters

Configure BFD sessions with default intervals and a default role.

The parameters that you can configure are: `desired tx interval`, `required min rx interval`, `detection multiplier`, and system `role`. Configure these parameters for all OSPFv3 sessions or all OSPFv3 sessions on a particular interface. If you change a parameter globally, the change affects all OSPFv3 neighbors sessions. If you change a parameter at the interface level, the change affects all OSPFv3 sessions on that interface.

To change parameters for all OSPFv3 sessions or for OSPFv3 sessions on a single interface, use the following commands.

To view session parameters, use the `show bfd neighbors detail` command, as shown in the example in Displaying BFD for BGP Information.

- Change parameters for all OSPFv3 sessions.

  ROUTER-OSPFv3 mode

  ```
  bfd all-neighbors interval milliseconds min_rx milliseconds multiplier value
  role [active | passive]
  ```
- Change parameters for OSPFv3 sessions on a single interface.

  INTERFACE mode

  ```
  ipv6 ospf bfd all-neighbors interval milliseconds min_rx milliseconds
  multiplier value role [active | passive]
  ```

Disabling BFD for OSPFv3

If you disable BFD globally, all sessions are torn down and sessions on the remote system are placed in a Down state.
If you disable BFD on an interface, sessions on the interface are torn down and sessions on the remote system are placed in a Down state. Disabling BFD does not trigger a change in BFD clients; a final Admin Down packet is sent before the session is terminated.

To disable BFD sessions, use the following commands.

- Disable BFD sessions with all OSPFv3 neighbors.

  ROUTER-OSPFv3 mode

  ```
  no bfd all-neighbors
  ```
- Disable BFD sessions with OSPFv3 neighbors on a single interface.

  INTERFACE mode

  ```
  ipv6 ospf bfd all-neighbors disable
  ```

Configure BFD for IS-IS

BFD for IS-IS is supported on the S6000 platform.

When using BFD with IS-IS, the IS-IS protocol registers with the BFD manager on the RPM. BFD sessions are then established with all neighboring interfaces participating in IS-IS. If a neighboring interface fails, the BFD agent on the line card notifies the BFD manager, which in turn notifies the IS-IS protocol that a link state change occurred.

Configuring BFD for IS-IS is a two-step process:

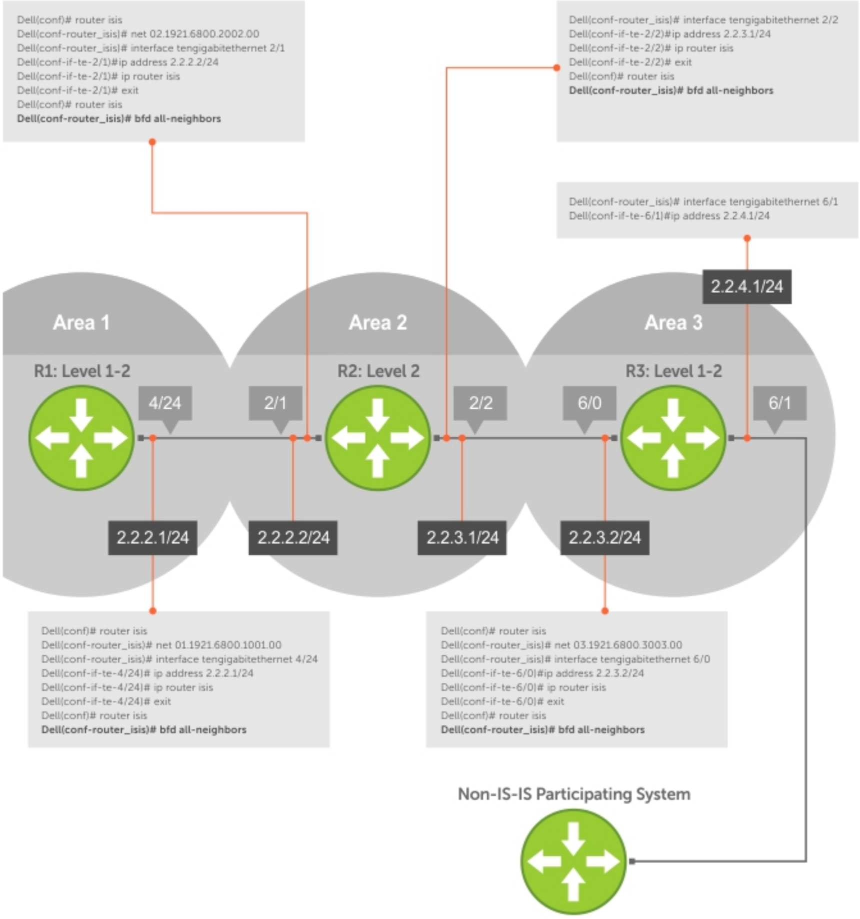
1. Enable BFD globally.
2. Establish sessions for all or particular IS-IS neighbors.

Related Configuration Tasks

- Changing IS-IS Session Parameters
- Disabling BFD for IS-IS

Establishing Sessions with IS-IS Neighbors

BFD sessions can be established for all IS-IS neighbors at once or sessions can be established for all neighbors out of a specific interface.



```
Dell(conf)# router isis
Dell(conf-router_isis)# net 02.1921.6800.2002.00
Dell(conf-router_isis)# interface tengigabitethernet 2/1
Dell(conf-if-te-2/1)#ip address 2.2.2.2/24
Dell(conf-if-te-2/1)# ip router isis
Dell(conf-if-te-2/1)# exit
Dell(conf)# router isis
Dell(conf-router_isis)# bfd all-neighbors
```

```
Dell(conf-router_isis)# interface tengigabitethernet 2/2
Dell(conf-if-te-2/2)#ip address 2.2.3.1/24
Dell(conf-if-te-2/2)# ip router isis
Dell(conf-if-te-2/2)# exit
Dell(conf)# router isis
Dell(conf-router_isis)# bfd all-neighbors
```

```
Dell(conf-router_isis)# interface tengigabitethernet 6/1
Dell(conf-if-te-6/1)#ip address 2.2.4.1/24
```

```
Dell(conf)# router isis
Dell(conf-router_isis)# net 01.1921.6800.1001.00
Dell(conf-router_isis)# interface tengigabitethernet 4/24
Dell(conf-if-te-4/24)# ip address 2.2.2.1/24
Dell(conf-if-te-4/24)# ip router isis
Dell(conf-if-te-4/24)# exit
Dell(conf)# router isis
Dell(conf-router_isis)# bfd all-neighbors
```

```
Dell(conf)# router isis
Dell(conf-router_isis)# net 03.1921.6800.3003.00
Dell(conf-router_isis)# interface tengigabitethernet 6/0
Dell(conf-if-te-6/0)#ip address 2.2.3.2/24
Dell(conf-if-te-6/0)# ip router isis
Dell(conf-if-te-6/0)# exit
Dell(conf)# router isis
Dell(conf-router_isis)# bfd all-neighbors
```

Figure 14. Establishing Sessions with IS-IS Neighbors

To establish BFD with all IS-IS neighbors or with IS-IS neighbors on a single interface, use the following commands.

- Establish sessions with all IS-IS neighbors.

  ROUTER-ISIS mode

  ```
  bfd all-neighbors
  ```
- Establish sessions with IS-IS neighbors on a single interface.

  INTERFACE mode

  ```
  isis bfd all-neighbors
  ```

Example of Verifying Sessions with IS-IS Neighbors

To view the established sessions, use the `show bfd neighbors` command.

The bold line shows that IS-IS BFD sessions are enabled.

```
R2(conf-router_isis)#bfd all-neighbors
R2(conf-router_isis)#do show bfd neighbors

*     - Active session role
Ad Dn - Admin Down
C     - CLI
I   - ISIS
O     - OSPF
R     - Static Route (RTM)

LocalAddr  RemoteAddr Interface State Rx-int Tx-int Mult Clients
* 2.2.2.2  2.2.2.1    Gi 2/1    Up    100    100    3    I
```

Changing IS-IS Session Parameters

BFD sessions are configured with default intervals and a default role.
The parameters that you can configure are: Desired TX Interval, Required Min RX Interval, Detection Multiplier, and system role. These parameters are configured for all IS-IS sessions or all IS-IS sessions out of an interface. If you change a parameter globally, the change affects all IS-IS neighbors sessions. If you change a parameter at the interface level, the change affects all IS-IS sessions on that interface.

To change parameters for all IS-IS sessions or for IS-IS sessions on a single interface, use the following commands.

To view session parameters, use the `show bfd neighbors detail` command, as shown in *Verifying BFD Sessions with BGP Neighbors Using the `show bfd neighbors` Command* in Displaying BFD for BGP Information.

- Change parameters for all IS-IS sessions.

  ROUTER-ISIS mode

  ```
  bfd all-neighbors interval milliseconds min_rx milliseconds multiplier value
  role [active | passive]
  ```
- Change parameters for IS-IS sessions on a single interface.

  INTERFACE mode

  ```
  isis bfd all-neighbors interval milliseconds min_rx milliseconds multiplier
  value role [active | passive]
  ```

Disabling BFD for IS-IS

If you disable BFD globally, all sessions are torn down and sessions on the remote system are placed in a Down state.
If you disable BFD on an interface, sessions on the interface are torn down and sessions on the remote system are placed in a Down state. Disabling BFD does not trigger a change in BFD clients; a final Admin Down packet is sent before the session is terminated.

To disable BFD sessions, use the following commands.

- Disable BFD sessions with all IS-IS neighbors.

  ROUTER-ISIS mode

  ```
  no bfd all-neighbors
  ```
- Disable BFD sessions with IS-IS neighbors on a single interface.

INTERFACE mose

```
isis bfd all-neighbors disable
```

Configure BFD for BGP

Bidirectional forwarding detection (BFD) for BGP is supported on the S6000 platform.

In a BGP core network, BFD provides rapid detection of communication failures in BGP fast-forwarding paths between internal BGP (iBGP) and external BGP (eBGP) peers for faster network reconvergence. BFD for BGP is supported on 1GE, 10GE, 40GE, port-channel, and VLAN interfaces. BFD for BGP does not support IPv6 and the BGP multihop feature.

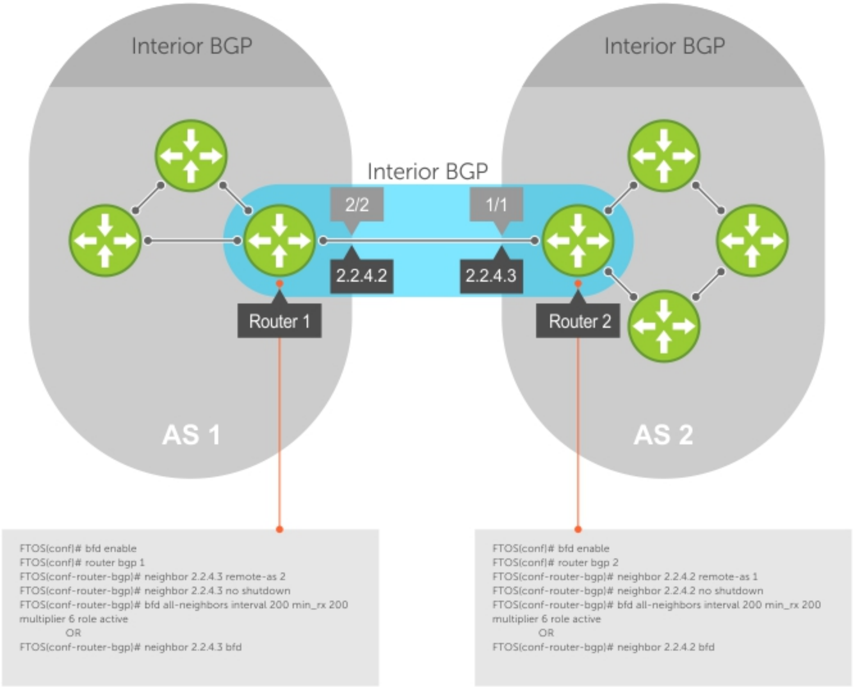
Prerequisites

Before configuring BFD for BGP, you must first configure the following settings:

1. Configure BGP on the routers that you want to interconnect, as described in Border Gateway Protocol IPv4 (BGPv4).
2. Enable fast fall-over for BGP neighbors to reduce convergence time (the `neighbor fall-over` command), as described in BGP Fast Fall-Over.

Establishing Sessions with BGP Neighbors

Before configuring BFD for BGP, you must first configure BGP on the routers that you want to interconnect.
For more information, refer to Border Gateway Protocol IPv4 (BGPv4).
For example, the following illustration shows a sample BFD configuration on Router 1 and Router 2 that use eBGP in a transit network to interconnect AS1 and AS2. The eBGP routers exchange information with each other as well as with iBGP routers to maintain connectivity and accessibility within each autonomous system.



```
FTOS(conf)# bfd enable
FTOS(conf)# router bgp 1
FTOS(conf-router-bgp)# neighbor 2.2.4.3 remote-as 2
FTOS(conf-router-bgp)# neighbor 2.2.4.3 no shutdown
FTOS(conf-router-bgp)# bfd all-neighbors interval 200 min_rx 200
multiplier 6 role active
         OR
FTOS(conf-router-bgp)# neighbor 2.2.4.3 bfd
```

```
FTOS(conf)# bfd enable
FTOS(conf)# router bgp 2
FTOS(conf-router-bgp)# neighbor 2.2.4.2 remote-as 1
FTOS(conf-router-bgp)# neighbor 2.2.4.2 no shutdown
FTOS(conf-router-bgp)# bfd all-neighbors interval 200 min_rx 200
multiplier 6 role active
         OR
FTOS(conf-router-bgp)# neighbor 2.2.4.2 bfd
```

Figure 15. Establishing Sessions with BGP Neighbors

The sample configuration shows alternative ways to establish a BFD session with a BGP neighbor:

- By establishing BFD sessions with all neighbors discovered by BGP (the `bfd all-neighbors` command).
- By establishing a BFD session with a specified BGP neighbor (the `neighbor {`*`ip-address`*` | `*`peer-group-name`*`} bfd` command)

BFD packets originating from a router are assigned to the highest priority egress queue to minimize transmission delays. Incoming BFD control packets received from the BGP neighbor are assigned to the highest priority queue within the control plane policing (COPP) framework to avoid BFD packets drops due to queue congestion.

BFD notifies BGP of any failure conditions that it detects on the link. Recovery actions are initiated by BGP.

BFD for BGP is supported only on directly-connected BGP neighbors and only in BGP IPv4 networks. Up to 128 simultaneous BFD sessions are supported

As long as each BFD for BGP neighbor receives a BFD control packet within the configured BFD interval for failure detection, the BFD session remains up and BGP maintains its adjacencies. If a BFD for BGP neighbor does not receive a control packet within the detection interval, the router informs any clients of the BFD session (other routing protocols) about the failure. It then depends on the individual routing protocols that uses the BGP link to determine the appropriate response to the failure condition. The typical response is to terminate the peering session for the routing protocol and reconverge by bypassing the failed neighboring router. A log message is generated whenever BFD detects a failure condition.

1. Enable BFD globally.

   CONFIGURATION mode

   ```
   bfd enable
   ```
2. Specify the AS number and enter ROUTER BGP configuration mode.

   CONFIGURATION mode

   ```
   router bgp as-number
   ```
3. Add a BGP neighbor or peer group in a remote AS.

   CONFIG-ROUTERBGP mode

   ```
   neighbor {ip-address | peer-group name} remote-as as-number
   ```
4. Enable the BGP neighbor.

   CONFIG-ROUTERBGP mode

   ```
   neighbor {ip-address | peer-group-name} no shutdown
   ```
5. Configure parameters for a BFD session established with all neighbors discovered by BGP. OR Establish a BFD session with a specified BGP neighbor or peer group using the default BFD session parameters.

   CONFIG-ROUTERBGP mode

   ```
   bfd all-neighbors [interval millisecs min_rx millisecs multiplier value role
   {active | passive}]
   ```

   OR

   ```
   neighbor {ip-address | peer-group-name} bfd
   ```

   NOTES:

   - When you establish a BFD session with a specified BGP neighbor or peer group using the `neighbor bfd` command, the default BFD session parameters are used (interval: 100 milliseconds, min_rx: 100 milliseconds, multiplier: 3 packets, and role: active).
   - When you explicitly enable or disable a BGP neighbor for a BFD session with the `neighbor bfd` or `neighbor bfd disable` commands, the neighbor does not inherit the BFD enable/disable values configured with the `bfd all-neighbors` command or configured for the peer group to which the neighbor belongs. Also, the neighbor only inherits the global timer values configured with the `bfd all-neighbors` command (interval, min_rx, and multiplier).
6. Repeat Steps 1 to 5 on each BGP peer participating in a BFD session.

Disabling BFD for BGP

You can disable BFD for BGP.
To disable a BFD for BGP session with a specified neighbor, use the first command. To remove the disabled state of a BFD for BGP session with a specified neighbor, use the second command.

The BGP link with the neighbor returns to normal operation and uses the BFD session parameters globally configured with the `bfd all-neighbors` command or configured for the peer group to which the neighbor belongs.

- Disable a BFD for BGP session with a specified neighbor.

ROUTER BGP mode

```
neighbor {ip-address | peer-group-name} bfd disable
```

- Remove the disabled state of a BFD for BGP session with a specified neighbor.

  ROUTER BGP mode

  ```
  no neighbor {ip-address | peer-group-name} bfd disable
  ```

Use BFD in a BGP Peer Group

You can establish a BFD session for the members of a peer group (the `neighbor peer-group-name bfd` command in ROUTER BGP configuration mode).

Members of the peer group may have BFD:

- Explicitly enabled (the `neighbor ip-address bfd` command)
- Explicitly disabled (the `neighbor ip-address bfd disable` command)
- Inherited (neither explicitly enabled or disabled) according to the current BFD configuration of the peer group. For information about BGP peer groups, refer to Configure Peer Groups.

If you explicitly enable (or disable) a BGP neighbor for BFD that belongs to a peer group:

- The neighbor does not inherit the BFD enable/disable values configured with the `bfd all-neighbors` command or configured for the peer group to which the neighbor belongs.
- The neighbor inherits only the global timer values that are configured with the `bfd all-neighbors` command (interval, min_rx, and multiplier).

If you explicitly enable (or disable) a peer group for BFD that has no BFD parameters configured (for example, advertisement interval) using the `neighbor` *`peer-group-name`* `bfd` command, the peer group inherits any BFD settings configured with the `bfd all-neighbors` command.

Displaying BFD for BGP Information

You can display related information for BFD for BGP.
To display information about BFD for BGP sessions on a router, use the following commands and refer to the following examples.

- Verify a BFD for BGP configuration.

  EXEC Privilege mode

  ```
  show running-config bgp
  ```

- Verify that a BFD for BGP session has been successfully established with a BGP neighbor. A line-by-line listing of established BFD adjacencies is displayed.

  EXEC Privilege mode

  ```
  show bfd neighbors [interface] [detail]
  ```

- Check to see if BFD is enabled for BGP connections.

  EXEC Privilege mode

  ```
  show ip bgp summary
  ```

- Displays routing information exchanged with BGP neighbors, including BFD for BGP sessions.

  EXEC Privilege mode

  ```
  show ip bgp neighbors [ip-address]
  ```

Examples of the BFD `show` Commands

The following example shows verifying a BGP configuration.

```
R2# show running-config bgp
!
router bgp 2
   neighbor 1.1.1.2 remote-as 1
   neighbor 1.1.1.2 no shutdown
   neighbor 2.2.2.2 remote-as 1
   neighbor 2.2.2.2 no shutdown
   neighbor 3.3.3.2 remote-as 1
   neighbor 3.3.3.2 no shutdown
   bfd all-neighbors
```

The following example shows viewing all BFD neighbors.

```
R2# show bfd neighbors

*     - Active session role
Ad Dn - Admin Down
B     - BGP
C     - CLI
I     - ISIS
O     - OSPF
R     - Static Route (RTM)
M     - MPLS
V     - VRRP

LocalAddr  RemoteAddr Interface State Rx-int Tx-int Mult Clients
* 1.1.1.3  1.1.1.2    Te 6/0    Up    100    100    3    B
* 2.2.2.3  2.2.2.2    Te 6/1    Up    100    100    3    B
* 3.3.3.3  3.3.3.2    Te 6/2    Up    100    100    3    B
```

The following example shows viewing BFD neighbors with full detail.

The bold lines show the BFD session parameters: TX (packet transmission), RX (packet reception), and multiplier (maximum number of missed packets).

```
R2# show bfd neighbors detail

Session Discriminator: 9
Neighbor Discriminator: 10
Local Addr: 1.1.1.3
Local MAC Addr: 00:01:e8:66:da:33
Remote Addr: 1.1.1.2
Remote MAC Addr: 00:01:e8:8a:da:7b
Int: TenGigabitEthernet 6/0
State: Up
Configured parameters:
TX: 100ms, RX: 100ms, Multiplier: 3
Neighbor parameters:
TX: 100ms, RX: 100ms, Multiplier: 3
Actual parameters:
TX: 100ms, RX: 100ms, Multiplier: 3
Role: Active
Delete session on Down: True
Client Registered: BGP
Uptime: 00:07:55
Statistics:
Number of packets received from neighbor: 4762
Number of packets sent to neighbor: 4490
Number of state changes: 2
Number of messages from IFA about port state change: 0
Number of messages communicated b/w Manager and Agent: 5

Session Discriminator: 10
Neighbor Discriminator: 11
Local Addr: 2.2.2.3
Local MAC Addr: 00:01:e8:66:da:34
Remote Addr: 2.2.2.2
Remote MAC Addr: 00:01:e8:8a:da:7b
Int: TenGigabitEthernet 6/1
State: Up
Configured parameters:
TX: 100ms, RX: 100ms, Multiplier: 3
Neighbor parameters:
TX: 100ms, RX: 100ms, Multiplier: 3
Actual parameters:
TX: 100ms, RX: 100ms, Multiplier: 3
Role: Active
Delete session on Down: True
Client Registered: BGP
Uptime: 00:02:22
Statistics:
Number of packets received from neighbor: 1428
Number of packets sent to neighbor: 1428
Number of state changes: 1
Number of messages from IFA about port state change: 0
Number of messages communicated b/w Manager and Agent: 4
```

The following example shows viewing configured BFD counters.

```
R2# show bfd counters bgp

Interface TenGigabitEthernet 6/0

Protocol BGP
Messages:
Registration    : 5
De-registration : 4
Init            : 0
Up              : 6
Down            : 0
Admin Down      : 2

Interface TenGigabitEthernet 6/1

Protocol BGP
Messages:
Registration    : 5
De-registration : 4
Init            : 0
Up              : 6
Down            : 0
Admin Down      : 2

Interface TenGigabitEthernet 6/2

Protocol BGP
Messages:
Registration    : 1
De-registration : 0
Init            : 0
Up              : 1
```

```
Down            : 0
Admin Down      : 2
```

The following example shows viewing BFD summary information.

The bold line shows the message displayed when you enable BFD for BGP connections.

```
R2# show ip bgp summary
BGP router identifier 10.0.0.1, local AS number 2
BGP table version is 0, main routing table version 0
```

BFD is enabled, Interval 100 Min_rx 100 Multiplier 3 Role Active

```
3 neighbor(s) using 24168 bytes of memory

Neighbor AS MsgRcvd MsgSent TblVer InQ OutQ Up/Down   State/Pfx

1.1.1.2  1  282     281     0      0   0    00:38:12  0
2.2.2.2  1  273     273     0      0  (0)   04:32:26  0
3.3.3.2  1  282     281     0      0   0    00:38:12  0
```

The following example shows viewing BFD information for a specified neighbor.

The bold lines show the message displayed when you enable a BFD session with different configurations:

- Message displays when you enable a BFD session with a BGP neighbor that inherits the global BFD session settings configured with the `global bfd all-neighbors` command.
- Message displays when you enable a BFD session with a BGP neighbor using the `neighbor` *`ip-address`* `bfd` command.
- Message displays when you enable a BGP neighbor in a peer group for which you enabled a BFD session using the `neighbor` *`peer-group-name`* `bfd` command

```
R2# show ip bgp neighbors 2.2.2.2

BGP neighbor is 2.2.2.2, remote AS 1, external link
  BGP version 4, remote router ID 12.0.0.4
  BGP state ESTABLISHED, in this state for 00:05:33
  Last read 00:00:30, last write 00:00:30
  Hold time is 180, keepalive interval is 60 seconds
  Received 8 messages, 0 in queue
    1 opens, 0 notifications, 0 updates
    7 keepalives, 0 route refresh requests
  Sent 9 messages, 0 in queue
    2 opens, 0 notifications, 0 updates
    7 keepalives, 0 route refresh requests
  Minimum time between advertisement runs is 30 seconds
  Minimum time before advertisements start is 0 seconds

Capabilities received from neighbor for IPv4 Unicast :
  MULTIPROTO_EXT(1)
  ROUTE_REFRESH(2)
  CISCO_ROUTE_REFRESH(128)

Capabilities advertised to neighbor for IPv4 Unicast :
  MULTIPROTO_EXT(1)
  ROUTE_REFRESH(2)
  CISCO_ROUTE_REFRESH(128)
```

Neighbor is using BGP global mode BFD configuration

```
For address family: IPv4 Unicast
BGP table version 0, neighbor version 0
Prefixes accepted 0 (consume 0 bytes), withdrawn 0 by peer, martian prefixes
ignored 0
Prefixes advertised 0, denied 0, withdrawn 0 from peer
```

```
  Connections established 1; dropped 0
  Last reset never
Local host: 2.2.2.3, Local port: 63805
Foreign host: 2.2.2.2, Foreign port: 179
E1200i_ExaScale#

R2# show ip bgp neighbors 2.2.2.3

BGP neighbor is 2.2.2.3, remote AS 1, external link
  Member of peer-group pg1 for session parameters
  BGP version 4, remote router ID 12.0.0.4
  BGP state ESTABLISHED, in this state for 00:05:33
  ...
  Neighbor is using BGP neighbor mode BFD configuration
  Peer active in peer-group outbound optimization
...

R2# show ip bgp neighbors 2.2.2.4

BGP neighbor is 2.2.2.4, remote AS 1, external link
  Member of peer-group pg1 for session parameters
  BGP version 4, remote router ID 12.0.0.4
  BGP state ESTABLISHED, in this state for 00:05:33
  ...
  Neighbor is using BGP peer-group mode BFD configuration
  Peer active in peer-group outbound optimization
  ...
```

Configure BFD for VRRP

BFD for VRRP is supported on the platform.

When using BFD with VRRP, the VRRP protocol registers with the BFD manager on the route processor module (RPM). BFD sessions are established with all neighboring interfaces participating in VRRP. If a neighboring interface fails, the BFD agent on the line card notifies the BFD manager, which in turn notifies the VRRP protocol that a link state change occurred.

Configuring BFD for VRRP is a three-step process:

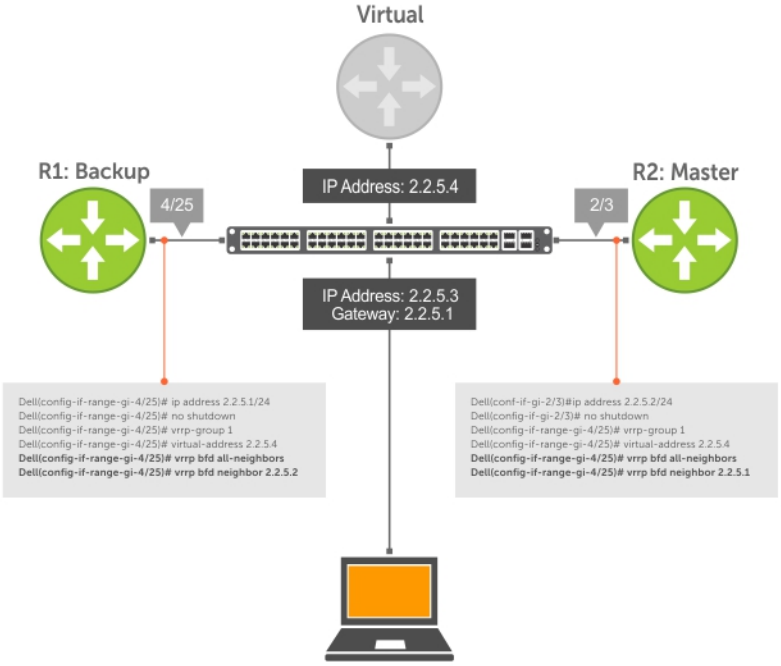
1. Enable BFD globally. Refer to Enabling BFD Globally.
2. Establish VRRP BFD sessions with all VRRP-participating neighbors.
3. On the master router, establish a VRRP BFD sessions with the backup routers. Refer to Establishing Sessions with All VRRP Neighbors.

Related Configuration Tasks

- Changing VRRP Session Parameters.
- Disabling BFD for VRRP.

Establishing Sessions with All VRRP Neighbors

BFD sessions can be established for all VRRP neighbors at once, or a session can be established with a particular neighbor.



```
Dell(config-if-range-gi-4/25)# ip address 2.2.5.1/24
Dell(config-if-range-gi-4/25)# no shutdown
Dell(config-if-range-gi-4/25)# vrrp-group 1
Dell(config-if-range-gi-4/25)# virtual-address 2.2.5.4
Dell(config-if-range-gi-4/25)# vrrp bfd all-neighbors
Dell(config-if-range-gi-4/25)# vrrp bfd neighbor 2.2.5.2
```

```
Dell(conf-if-gi-2/3)#ip address 2.2.5.2/24
Dell(config-if-gi-2/3)# no shutdown
Dell(config-if-range-gi-4/25)# vrrp-group 1
Dell(config-if-range-gi-4/25)# virtual-address 2.2.5.4
Dell(config-if-range-gi-4/25)# vrrp bfd all-neighbors
Dell(config-if-range-gi-4/25)# vrrp bfd neighbor 2.2.5.1
```

Figure 16. Establishing Sessions with All VRRP Neighbors

To establish sessions with all VRRP neighbors, use the following command.

- Establish sessions with all VRRP neighbors.

  INTERFACE mode

  ```
  vrrp bfd all-neighbors
  ```

Establishing VRRP Sessions on VRRP Neighbors

The master router does not care about the state of the backup router, so it does not participate in any VRRP BFD sessions.
VRRP BFD sessions on the backup router cannot change to the UP state. Configure the master router to establish an individual VRRP session the backup router.

To establish a session with a particular VRRP neighbor, use the following command.

- Establish a session with a particular VRRP neighbor.

  INTERFACE mode

  ```
  vrrp bfd neighbor ip-address
  ```

Examples of Viewing VRRP Sessions with Neighbors or State Information

To view the established sessions, use the `show bfd neighbors` command.

The bold line shows that VRRP BFD sessions are enabled.

```
Dell(conf-if-gi-4/25)#vrrp bfd all-neighbors
Dell(conf-if-gi-4/25)#do show bfd neighbor

*      - Active session role
Ad Dn - Admin Down
C      - CLI
I      - ISIS
O      - OSPF
R      - Static Route (RTM)
V      - VRRP

LocalAddr  RemoteAddr Interface State Rx-int Tx-int Mult Clients
* 2.2.5.1  2.2.5.2    Gi 4/25  Down 1000   1000   3   V
```

To view session state information, use the `show vrrp` command.

The bold line shows the VRRP BFD session.

```
Dell(conf-if-gi-4/25)#do show vrrp
------------------
GigabitEthernet 4/1, VRID: 1, Net: 2.2.5.1
State: Backup, Priority: 1, Master: 2.2.5.2
Hold Down: 0 sec, Preempt: TRUE, AdvInt: 1 sec
Adv rcvd: 95, Bad pkts rcvd: 0, Adv sent: 933, Gratuitous ARP sent: 3
Virtual MAC address:
  00:00:5e:00:01:01
Virtual IP address:
  2.2.5.4
Authentication: (none)
BFD Neighbors:
RemoteAddr  State
2.2.5.2     Up
```

Changing VRRP Session Parameters

BFD sessions are configured with default intervals and a default role.
The parameters that you can configure are: Desired TX Interval, Required Min RX Interval, Detection Multiplier, and system role. You can change parameters for all VRRP sessions or for a particular neighbor.

To change parameters for all VRRP sessions or for a particular VRRP session, use the following commands.

- Change parameters for all VRRP sessions.

  INTERFACE mode

  ```
  vrrp bfd all-neighbors interval milliseconds min_rx milliseconds multiplier
  value role [active | passive]
  ```

- Change parameters for a particular VRRP session.

  INTERFACE mode

  ```
  vrrp bfd neighbor ip-address interval milliseconds min_rx milliseconds
  multiplier value role [active | passive]
  ```

To view session parameters, use the `show bfd neighbors detail` command, as shown in the example in *Verifying BFD Sessions with BGP Neighbors Using the* `show bfd neighbors` *command* example in Displaying BFD for BGP Information.

Disabling BFD for VRRP

If you disable any or all VRRP sessions, the sessions are torn down.
A final Admin Down control packet is sent to all neighbors and sessions on the remote system change to the Down state.

To disable all VRRP sessions on an interface, sessions for a particular VRRP group, or for a particular VRRP session on an interface, use the following commands.

- Disable all VRRP sessions on an interface.

  INTERFACE mode

  ```
  no vrrp bfd all-neighbors
  ```
- Disable all VRRP sessions in a VRRP group.

  VRRP mode

  ```
  bfd disable
  ```
- Disable a particular VRRP session on an interface.

  INTERFACE mode

  ```
  no vrrp bfd neighbor ip-address
  ```

Configuring Protocol Liveness

Protocol liveness is a feature that notifies the BFD manager when a client protocol is disabled.
When you disable a client, all BFD sessions for that protocol are torn down. Neighbors on the remote system receive an Admin Down control packet and are placed in the Down state.

To enable protocol liveness, use the following command.

- Enable Protocol Liveness.

  CONFIGURATION mode

  ```
  bfd protocol-liveness
  ```

Troubleshooting BFD

To troubleshoot BFD, use the following commands and examples.
To control packet field values or to examine the control packets in hexadecimal format, use the following command.

- Examine control packet field values.

  CONFIGURATION mode

  ```
  debug bfd detail
  ```
- Examine the control packets in hexadecimal format.

  CONFIGURATION

  ```
  debug bfd packet
  ```

Examples of Output from the `debug bfd` Commands

The following example shows a three-way handshake using the `debug bfd detail` command.

```
     R1(conf-if-gi-4/24)#00:54:38: %RPM0-P:RP2 %BFDMGR-1-BFD_STATE_CHANGE:
Changed session state to
```

```
Down for neighbor 2.2.2.2 on interface Gi 4/24 (diag: 0)
     00:54:38 : Sent packet for session with neighbor 2.2.2.2 on Gi 4/24
     TX packet dump:
       Version:1, Diag code:0, State:Down, Poll bit:0, Final bit:0, Demand bit:0
       myDiscrim:4, yourDiscrim:0, minTx:1000000, minRx:1000000, multiplier:3,
minEchoRx:0
     00:54:38 : Received packet for session with neighbor 2.2.2.2 on Gi 4/24
     RX packet dump:
       Version:1, Diag code:0, State:Init, Poll bit:0, Final bit:0, Demand bit:0
       myDiscrim:6, yourDiscrim:4, minTx:1000000, minRx:1000000, multiplier:3,
minEchoRx:0
     00:54:38: %RPM0-P:RP2 %BFDMGR-1-BFD_STATE_CHANGE: Changed session state to Up
for neighbor 2.2.2.2
on interface Gi 4/24 (diag: 0)
```

The following example shows hexadecimal output from the `debug bfd packet` command.

```
RX packet dump:
    20 c0 03 18 00 00 00 05 00 00 00 04 00 01 86 a0
    00 01 86 a0 00 00 00 00
    00:34:13 : Sent packet for session with neighbor 2.2.2.2 on Gi 4/24
TX packet dump:
    20 c0 03 18 00 00 00 04 00 00 00 05 00 01 86 a0
    00 01 86 a0 00 00 00 00
00:34:14 : Received packet for session with neighbor 2.2.2.2 on Gi 4/24
RX packet dump:
    20 c0 03 18 00 00 00 05 00 00 00 04 00 01 86 a0
    00 01 86 a0 00 00 00 00
00:34:14 : Sent packet for session with neighbor 2.2.2.2 on Gi 4/24
TX packet dump:
    20 c0 03 18 00 00 00 04 00 00 00 05 00 01 86 a0
    00 01 86 a0 00 00 00 00
00:34:14 : Received packet for session with neighbor 2.2.2.2 on Gi 4/24
RX packet dump:
    20 c0 03 18 00 00 00 05 00 00 00 04 00 01 86 a0
    00 01 86 a0 00 00 00 00
00:34:14 : Sent packet for session with neighbor 2.2.2.2 on Gi 4/24
L
```

The output for the `debug bfd event` command is the same as the log messages that appear on the console by default.

8

Border Gateway Protocol IPv4 (BGPv4)

Border gateway protocol IPv4 (BGPv4) version 4 (BGPv4) is supported on the S6000 platform.

This chapter provides a general description of BGPv4 as it is supported in the Dell Networking Operating System (OS).

BGP protocol standards are listed in the Standards Compliance chapter.

BGP is an external gateway protocol that transmits interdomain routing information within and between autonomous systems (AS). The primary function of the BGP is to exchange network reachability information with other BGP systems. BGP generally operates with an internal gateway protocol (IGP) such as open shortest path first (OSPF) or router information protocol (RIP), allowing you to communicate to external ASs smoothly. BGP adds reliability to network connections by having multiple paths from one router to another.

Autonomous Systems (AS)

BGP autonomous systems (ASs) are a collection of nodes under common administration with common network routing policies.

Each AS has a number, which an internet authority already assigns. You do not assign the BGP number.

AS numbers (ASNs) are important because the ASN uniquely identifies each network on the internet. The Internet Assigned Numbers Authority (IANA) has reserved AS numbers 64512 through 65534 to be used for private purposes. IANA reserves ASNs 0 and 65535 and must not be used in a live environment.

You can group autonomous systems into three categories (multihomed, stub, and transit), defined by their connections and operation.

- **multihomed AS** — is one that maintains connections to more than one other AS. This group allows the AS to remain connected to the Internet in the event of a complete failure of one of their connections. However, this type of AS does not allow traffic from one AS to pass through on its way to another AS. A simple example of this group is seen in the following illustration.
- **stub AS** — is one that is connected to only one other AS.
- **transit AS** — is one that provides connections through itself to separate networks. For example, in the following illustration, Router 1 can use Router 2 (the transit AS) to connect to Router 4. Internet service providers (ISPs) are always transit ASs, because they provide connections from one network to another. The ISP is considered to be “selling transit service” to the customer network, so thus the term Transit AS.

When BGP operates inside an AS (AS1 or AS2, as seen in the following illustration), it is referred to as Internal BGP (IBGP Internal Border Gateway Protocol). When BGP operates between ASs (AS1 and AS2), it is called External BGP (EBGP External Border Gateway Protocol). IBGP provides routers inside the AS with the knowledge to reach routers external to the AS. EBGP routers exchange information with other EBGP routers as well as IBGP routers to maintain connectivity and accessibility.

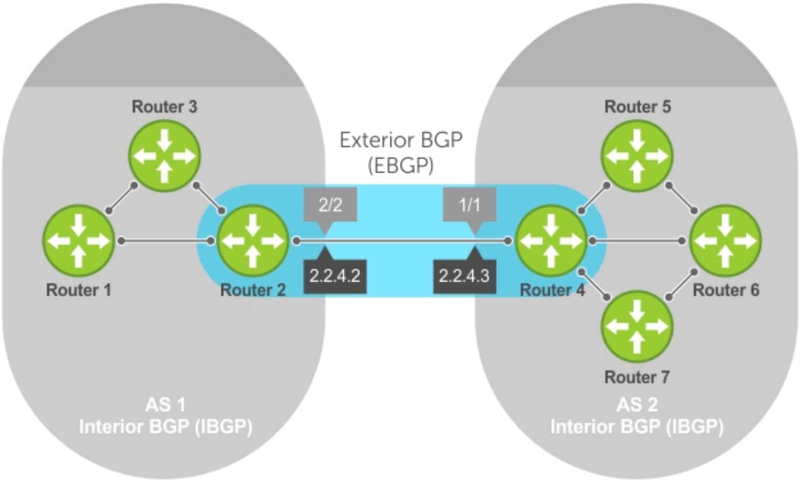

Router 3
Exterior BGP
(EBGP)
Router 5
2/2
1/1
Router 1
Router 2
2.2.4.2
2.2.4.3
Router 4
Router 6
Router 7
AS 1
Interior BGP (IBGP)
AS 2
Interior BGP (IBGP)

Figure 17. Internal BGP

BGP version 4 (BGPv4) supports classless interdomain routing and aggregate routes and AS paths. BGP is a path vector protocol — a computer network in which BGP maintains the path that updated information takes as it diffuses through the network. Updates traveling through the network and returning to the same node are easily detected and discarded.

BGP does not use a traditional interior gateway protocol (IGP) matrix, but makes routing decisions based on path, network policies, and/or rulesets. Unlike most protocols, BGP uses TCP as its transport protocol.

Since each BGP router talking to another router is a session, a BGP network needs to be in “full mesh.” This is a topology that has every router directly connected to every other router. Each BGP router within an AS must have iBGP sessions with all other BGP routers in the AS. For example, a BGP network within an AS needs to be in “full mesh.” As seen in the illustration below, four routers connected in a full mesh have three peers each, six routers have five peers each, and eight routers in full mesh have seven peers each.

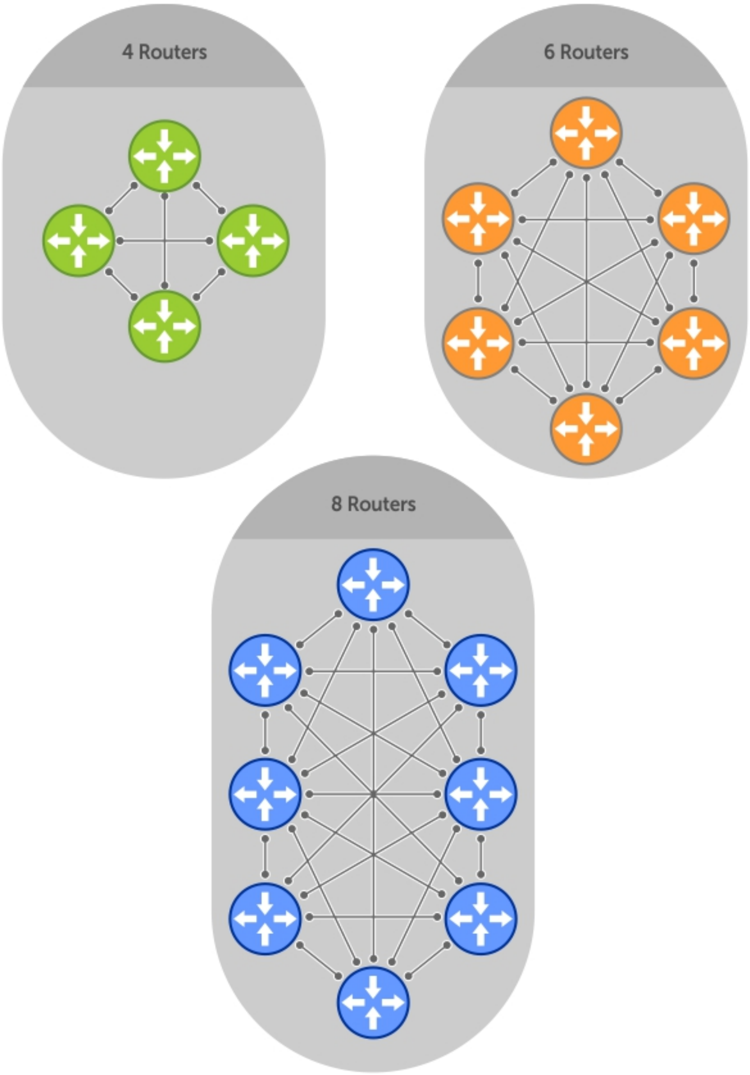


Figure 18. BGP Routers in Full Mesh

The number of BGP speakers each BGP peer must maintain increases exponentially. Network management quickly becomes impossible.

Sessions and Peers

When two routers communicate using the BGP protocol, a BGP session is started. The two end-points of that session are Peers. A Peer is also called a Neighbor.

Establish a Session

Information exchange between peers is driven by events and timers. The focus in BGP is on the traffic routing policies.

In order to make decisions in its operations with other BGP peers, a BGP process uses a simple finite state machine that consists of six states: Idle, Connect, Active, OpenSent, OpenConfirm, and Established. For each peer-to-peer session, a BGP implementation tracks which of these six states the session is in. The BGP protocol defines the messages that each peer should exchange in order to change the session from one state to another.

State	Description
Idle	BGP initializes all resources, refuses all inbound BGP connection attempts, and initiates a TCP connection to the peer.
Connect	In this state the router waits for the TCP connection to complete, transitioning to the OpenSent state if successful. If that transition is not successful, BGP resets the ConnectRetry timer and transitions to the Active state when the timer expires.
Active	The router resets the ConnectRetry timer to zero and returns to the Connect state.
OpenSent	After successful OpenSent transition, the router sends an Open message and waits for one in return.
OpenConfirm	After the Open message parameters are agreed between peers, the neighbor relation is established and is in the OpenConfirm state. This is when the router receives and checks for agreement on the parameters of open messages to establish a session.
Established	Keepalive messages are exchanged next, and after successful receipt, the router is placed in the Established state. Keepalive messages continue to be sent at regular periods (established by the Keepalive timer) to verify connections.

After the connection is established, the router can now send/receive Keepalive, Update, and Notification messages to/from its peer.

Peer Groups

Peer Ggroups are neighbors grouped according to common routing policies. They enable easier system configuration and management by allowing groups of routers to share and inherit policies.

Peer groups also aid in convergence speed. When a BGP process needs to send the same information to a large number of peers, the BGP process needs to set up a long output queue to get that information to all the proper peers. If the peers are members of a peer group however, the information can be sent to one place and then passed onto the peers within the group.

Route Reflectors

Route reflectors reorganize the iBGP core into a hierarchy and allow some route advertisement rules.

NOTE: Do not use route reflectors (RRs) in the forwarding path. In iBGP, hierarchal RRs maintaining forwarding plane RRs could create routing loops.

Route reflection divides iBGP peers into two groups: client peers and nonclient peers. A route reflector and its client peers form a route reflection cluster. Because BGP speakers announce only the best route for a given prefix, route reflector rules are applied after the router makes its best path decision.

- If a route was received from a nonclient peer, reflect the route to all client peers.
- If the route was received from a client peer, reflect the route to all nonclient and all client peers.

To illustrate how these rules affect routing, refer to the following illustration and the following steps. Routers B, C, D, E, and G are members of the same AS (AS100). These routers are also in the same Route Reflection Cluster, where Router D is the Route Reflector. Router E and H are client peers of Router D; Routers B and C and nonclient peers of Router D.

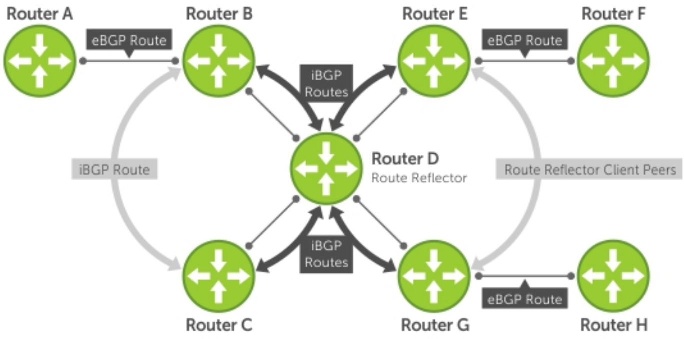


Figure 19. BGP Router Rules

1. Router B receives an advertisement from Router A through eBGP. Because the route is learned through eBGP, Router B advertises it to all its iBGP peers: Routers C and D.
2. Router C receives the advertisement but does not advertise it to any peer because its only other peer is Router D, an iBGP peer, and Router D has already learned it through iBGP from Router B.
3. Router D does not advertise the route to Router C because Router C is a nonclient peer and the route advertisement came from Router B who is also a nonclient peer.
4. Router D does reflect the advertisement to Routers E and G because they are client peers of Router D.
5. Routers E and G then advertise this iBGP learned route to their eBGP peers Routers F and H.

BGP Attributes

Routes learned using BGP have associated properties that are used to determine the best route to a destination when multiple paths exist to a particular destination.

These properties are referred to as BGP attributes, and an understanding of how BGP attributes influence route selection is required for the design of robust networks. This section describes the attributes that BGP uses in the route selection process:

- Weight
- Local Preference
- Multi-Exit Discriminators (MEDs)
- Origin
- AS Path

- Next Hop

NOTE: There are no hard coded limits on the number of attributes that are supported in the BGP. Taking into account other constraints such as the Packet Size, maximum number of attributes are supported in BGP.

Communities

BGP communities are sets of routes with one or more common attributes. Communities are a way to assign common attributes to multiple routes at the same time.

NOTE: Duplicate communities are not rejected.

Best Path Selection Criteria

Paths for active routes are grouped in ascending order according to their neighboring external AS number (BGP best path selection is deterministic by default, which means the `bgp non-deterministic-med` command is NOT applied).

The best path in each group is selected based on specific criteria. Only one “best path” is selected at a time. If any of the criteria results in more than one path, BGP moves on to the next option in the list. For example, two paths may have the same weights, but different local preferences. BGP sees that the Weight criteria results in two potential “best paths” and moves to local preference to reduce the options. If a number of best paths is determined, this selection criteria is applied to group’s best to determine the ultimate best path.

In non-deterministic mode (the `bgp non-deterministic-med` command is applied), paths are compared in the order in which they arrive. This method can lead to Dell Networking OS choosing different best paths from a set of paths, depending on the order in which they were received from the neighbors because MED may or may not get compared between the adjacent paths. In deterministic mode, Dell Networking OS compares MED between the adjacent paths within an AS group because all paths in the AS group are from the same AS.

NOTE: The `bgp bestpath as-path multipath-relax` command is disabled by default, preventing BGP from load-balancing a learned route across two or more eBGP peers. To enable load-balancing across different eBGP peers, enable the `bgp bestpath as-path multipath-relax` command. A system error results if you configure the `bgp bestpath as-path ignore` command and the `bgp bestpath as-path multipath-relax` command at the same time. Only enable one command at a time.

The following illustration shows that the decisions BGP goes through to select the best path. The list following the illustration details the path selection criteria.

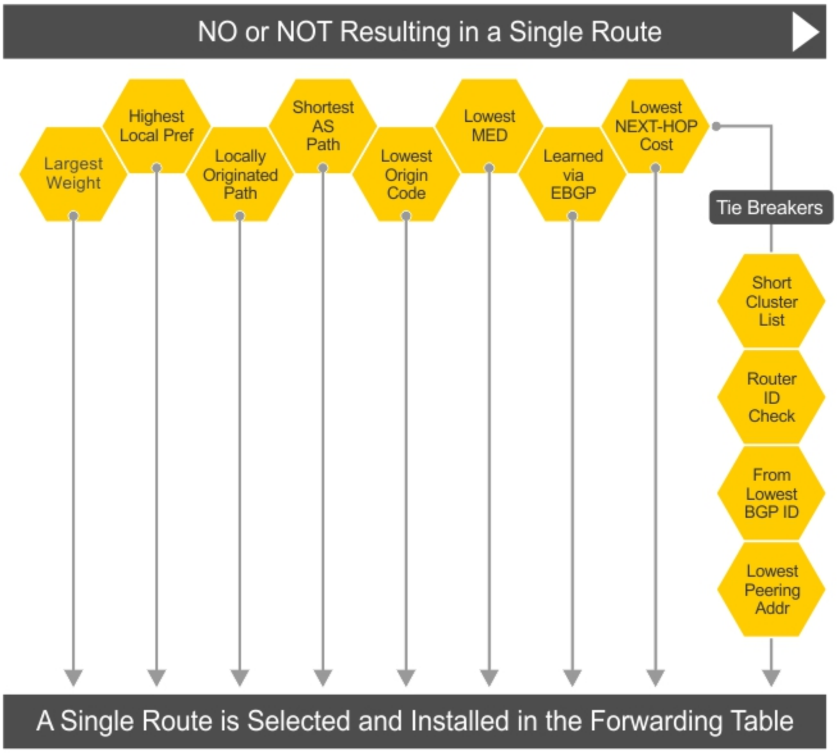

NO or NOT Resulting in a Single Route
Largest Weight
Highest Local Pref
Locally Originated Path
Shortest AS Path
Lowest Origin Code
Lowest MED
Learned via EBGP
Lowest NEXT-HOP Cost
Tie Breakers
Short Cluster List
Router ID Check
From Lowest BGP ID
Lowest Peering Addr
A Single Route is Selected and Installed in the Forwarding Table

Figure 20. BGP Best Path Selection

Best Path Selection Details

1. Prefer the path with the largest WEIGHT attribute.
2. Prefer the path with the largest LOCAL_PREF attribute.
3. Prefer the path that was locally `Originated via a network` command, `redistribute` command or `aggregate-address` command.
    a. Routes originated with the `Originated via a network` or `redistribute` commands are preferred over routes originated with the `aggregate-address` command.
4. Prefer the path with the shortest AS_PATH (unless the `bgp bestpath as-path ignore` command is configured, then AS_PATH is not considered). The following criteria apply:
    a. An AS_SET has a path length of 1, no matter how many ASs are in the set.
    b. A path with no AS_PATH configured has a path length of 0.
    c. AS_CONFED_SET is not included in the AS_PATH length.
    d. AS_CONFED_SEQUENCE has a path length of 1, no matter how many ASs are in the AS_CONFED_SEQUENCE.
5. Prefer the path with the lowest ORIGIN type (IGP is lower than EGP, and EGP is lower than INCOMPLETE).
6. Prefer the path with the lowest multi-exit discriminator (MED) attribute. The following criteria apply:
    a. This comparison is only done if the first (neighboring) AS is the same in the two paths; the MEDs are compared only if the first AS in the AS_SEQUENCE is the same for both paths.
    b. If you entered the `bgp always-compare-med` command, MEDs are compared for all paths.

    c. Paths with no MED are treated as “worst” and assigned a MED of 4294967295.
7. Prefer external (EBGP) to internal (IBGP) paths or confederation EBGP paths.
8. Prefer the path with the lowest IGP metric to the BGP if next-hop is selected when `synchronization` is disabled and only an internal path remains.
9. Dell Networking OS deems the paths as equal and does not perform steps 9 through 11, if the following criteria is met:
    a. the IBGP multipath or EBGP multipath are configured (the `maximum-path` command).
    b. the paths being compared were received from the same AS with the same number of ASs in the AS Path but with different NextHops.
    c. the paths were received from IBGP or EBGP neighbor respectively.
10. If the `bgp bestpath router-id ignore` command is enabled and:
    a. if the Router-ID is the same for multiple paths (because the routes were received from the same route) skip this step.
    b. if the Router-ID is NOT the same for multiple paths, prefer the path that was first received as the Best Path. The path selection algorithm returns without performing any of the checks detailed here.
11. Prefer the external path originated from the BGP router with the lowest router ID. If both paths are external, prefer the oldest path (first received path). For paths containing a route reflector (RR) attribute, the originator ID is substituted for the router ID.
12. If two paths have the same router ID, prefer the path with the lowest cluster ID length. Paths without a cluster ID length are set to a 0 cluster ID length.
13. Prefer the path originated from the neighbor with the lowest address. (The neighbor address is used in the BGP neighbor configuration and corresponds to the remote peer used in the TCP connection with the local router.)

After a number of best paths is determined, this selection criteria is applied to group’s best to determine the ultimate best path.

In non-deterministic mode (the `bgp non-deterministic-med` command is applied), paths are compared in the order in which they arrive. This method can lead to Dell Networking OS choosing different best paths from a set of paths, depending on the order in which they were received from the neighbors because MED may or may not get compared between the adjacent paths. In deterministic mode, Dell Networking OS compares MED between the adjacent paths within an AS group because all paths in the AS group are from the same AS.

Weight

The weight attribute is local to the router and is not advertised to neighboring routers.

If the router learns about more than one route to the same destination, the route with the highest weight is preferred. The route with the highest weight is installed in the IP routing table.

Local Preference

Local preference (LOCAL_PREF) represents the degree of preference within the entire AS. The higher the number, the greater the preference for the route.

Local preference (LOCAL_PREF) is one of the criteria used to determine the best path, so keep in mind that other criteria may impact selection, as shown in the illustration in Best Path Selection Criteria. For this example, assume that thelocal preference (LOCAL_PREF) is the only attribute applied. In the following illustration, AS100 has two possible paths to AS 200. Although the path through Router A is shorter (one hop instead of two), the LOCAL_PREF settings have the preferred path go through Router B and AS300. This is advertised to all routers within AS100, causing all BGP speakers to prefer the path through Router B.

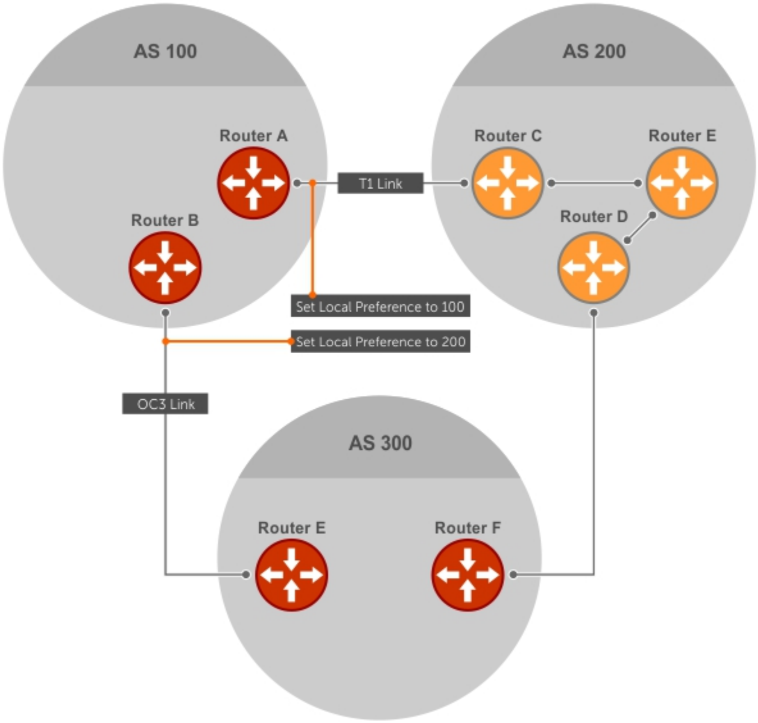

AS 100
AS 200
Router A
T1 Link
Router C
Router E
Router B
Router D
Set Local Preference to 100
Set Local Preference to 200
OC3 Link
AS 300
Router E
Router F

Figure 21. BGP Local Preference

Multi-Exit Discriminators (MEDs)

If two ASs connect in more than one place, a multi-exit discriminator (MED) can be used to assign a preference to a preferred path.

MED is one of the criteria used to determine the best path, so keep in mind that other criteria may impact selection, as shown in the illustration in Best Path Selection Criteria.

One AS assigns the MED a value and the other AS uses that value to decide the preferred path. For this example, assume the MED is the only attribute applied. In the following illustration, AS100 and AS200 connect in two places. Each connection is a BGP session. AS200 sets the MED for its T1 exit point to 100 and the MED for its OC3 exit point to 50. This sets up a path preference through the OC3 link. The MEDs are advertised to AS100 routers so they know which is the preferred path.

MEDs are non-transitive attributes. If AS100 sends an MED to AS200, AS200 does not pass it on to AS300 or AS400. The MED is a locally relevant attribute to the two participating ASs (AS100 and AS200).

NOTE: The MEDs are advertised across both links, so if a link goes down, AS 1 still has connectivity to AS300 and AS400.

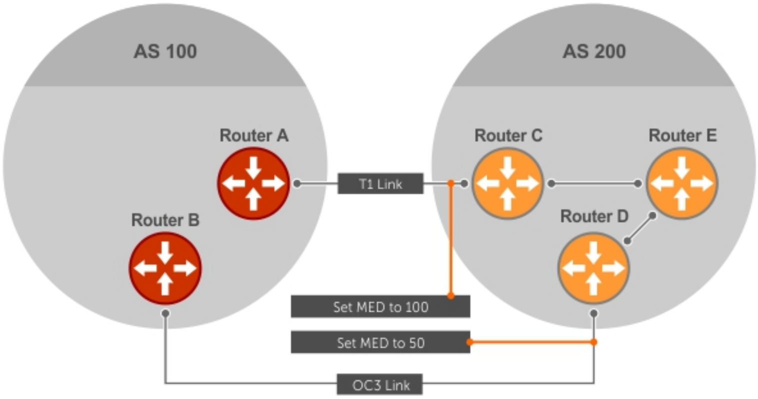


Figure 22. Multi-Exit Discriminators

NOTE: Configuring the `set metric-type internal` command in a route-map advertises the IGP cost as MED to outbound EBGP peers when redistributing routes. The configured `set metric` value overwrites the default IGP cost. If the outbound route-map uses MED, it overwrites IGP MED.

Origin

The origin indicates the origin of the prefix, or how the prefix came into BGP. There are three origin codes: IGP, EGP, INCOMPLETE.

Origin Type	Description
IGP	Indicates the prefix originated from information learned through an interior gateway protocol.
EGP	Indicates the prefix originated from information learned from an EGP protocol, which NGP replaced.
INCOMPLETE	Indicates that the prefix originated from an unknown source.

Generally, an IGP indicator means that the route was derived inside the originating AS. EGP generally means that a route was learned from an external gateway protocol. An INCOMPLETE origin code generally results from aggregation, redistribution, or other indirect ways of installing routes into BGP.

In Dell Networking OS, these origin codes appear as shown in the following example. The question mark (?) indicates an origin code of INCOMPLETE (shown in bold). The lower case letter (i) indicates an origin code of IGP (shown in bold).

Example of Viewing Origin Codes

```
Dell#show ip bgp
BGP table version is 0, local router ID is 10.101.15.13
Status codes: s suppressed, d damped, h history, * valid, > best
Path source: I - internal, a - aggregate, c - confed-external, r -
redistributed, n - network
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network      Next Hop      Metric  LocPrf  Weight  Path
*> 7.0.0.0/29  10.114.8.33  0        0       18508   ?
*> 7.0.0.0/30  10.114.8.33  0       0       18508   ?
*> 9.2.0.0/16  10.114.8.33  10      0       18508   701 i
```

AS Path

The AS path is the list of all ASs that all the prefixes listed in the update have passed through.

The local AS number is added by the BGP speaker when advertising to a eBGP neighbor.

NOTE: Any update that contains the AS path number 0 is valid.

The AS path is shown in the following example. The origin attribute is shown following the AS path information (shown in bold).

Example of Viewing AS Paths

```
Dell#show ip bgp paths
Total 30655 Paths
Address     Hash Refcount Metric Path
0x4014154   0    3        18508    701 3549 19421 i
0x4013914   0    3        18508    701 7018 14990 i
0x5166d6c   0    3        18508    209 4637 1221 9249 9249 i
0x5e62df4   0    2        18508    701 17302 i
0x3a1814c   0    26       18508    209 22291 i
0x567ea9c   0    75       18508    209 3356 2529 i
0x6cc1294   0    2        18508    209 1239 19265 i
0x6cc18d4   0    1        18508    701 2914 4713 17935 i
0x5982e44   0    162      18508    209 i
0x67d4a14   0    2        18508    701 19878 ?
0x559972c   0    31       18508    209 18756 i
0x59cd3b4   0    2        18508    209 7018 15227 i
0x7128114   0    10       18508    209 3356 13845 i
0x536a914   0    3        18508    209 701 6347 7781 i
0x2ffe884   0    1        18508    701 3561 9116 21350 i
```

Next Hop

The next hop is the IP address used to reach the advertising router.

For EBGP neighbors, the next-hop address is the IP address of the connection between the neighbors. For IBGP, the EBGP next-hop address is carried into the local AS. A next hop attribute is set when a BGP speaker advertises itself to another BGP speaker outside its local AS and when advertising routes within an AS. The next hop attribute also serves as a way to direct traffic to another BGP speaker, rather than waiting for a speaker to advertise. When a next-hop BGP neighbor is unreachable, then the connection to that BGP neighbor goes down after hold down timer expiry. The connection flap can also be obtained immediately with Fallover enabled. BGP routes that contain the next-hop as the neighbor address are not sent to the neighbor. You can enable this feature using the `neighbor sender-side-loopdetect` command.

NOTE: For EBGP neighbors, the next-hop address corresponding to a BGP route is not resolved if the next-hop address is not the same as the neighbor IP address.

NOTE: The connection between a router and its next-hop BGP neighbor terminates immediately only if the router has received routes from the BGP neighbor in the past.

Multiprotocol BGP

Multiprotocol extensions for BGP (MBGP) is defined in IETF RFC 2858. MBGP allows different types of address families to be distributed in parallel.

MBGP for IPv4 multicast is supported on the S6000 platform.

MBGP allows information about the topology of the IP multicast-capable routers to be exchanged separately from the topology of normal IPv4 and IPv6 unicast routers. It allows a multicast routing topology different from the unicast routing topology.

MBGP uses either an IPv4 address configured on the interface (which is used to establish the IPv6 session) or a stable IPv4 address that is available in the box as the next-hop address. As a result, while advertising an IPv6 network, exchange of IPv4 routes does not lead to martian next-hop message logs.

NOTE: It is possible to configure BGP peers that exchange both unicast and multicast network layer reachability information (NLRI), but you cannot connect multiprotocol BGP with BGP. Therefore, you cannot redistribute multiprotocol BGP routes into BGP.

Implement BGP with Dell Networking OS

The following sections describe how to implement BGP on Dell Networking OS.

Additional Path (Add-Path) Support

BGP add-path is supported on the platform.

The add-path feature reduces convergence times by advertising multiple paths to its peers for the same address prefix without replacing existing paths with new ones. By default, a BGP speaker advertises only the best path to its peers for a given address prefix. If the best path becomes unavailable, the BGP speaker withdraws its path from its local RIB and recalculates a new best path. This situation requires both IGP and BGP convergence and can be a lengthy process. BGP add-path also helps switchover to the next new best path when the current best path is unavailable.

Advertise IGP Cost as MED for Redistributed Routes

When using multipath connectivity to an external AS, you can advertise the MED value selectively to each peer for redistributed routes. For some peers you can set the internal/IGP cost as the MED while setting others to a constant pre-defined metric as MED value.

Dell Networking OS supports configuring the `set metric-type internal` command in a route-map to advertise the IGP cost as the MED to outbound EBGP peers when redistributing routes. The configured `set metric` value overwrites the default IGP cost.

By using the `redistribute` command with the `route-map` command, you can specify whether a peer advertises the standard MED or uses the IGP cost as the MED.

When configuring this functionality:

- If the `redistribute` command does not have `metric` configured and the BGP peer outbound route-map does have `metric-type internal` configured, BGP advertises the IGP cost as MED.
- If the `redistribute` command has `metric` configured (`route-map set metric` or `redistribute` *`route-type`* `metric`) and the BGP peer outbound route-map has `metric-type`

`internal` configured, BGP advertises the metric configured in the `redistribute` command as MED.
- If BGP peer outbound route-map has `metric` configured, all other metrics are overwritten by this configuration.

NOTE: When redistributing static, connected, or OSPF routes, there is no `metric` option. Simply assign the appropriate route-map to the redistributed route.

The following table lists some examples of these rules.

Table 6. Redistributed Route Rules

Command Settings	BGP Local Routing Information Base	MED Advertised to Peer WITH route-map metric-type internal	MED Advertised to Peer WITHOUT route-map metric-type internal
redistribute isis (IGP cost = 20)	MED: IGP cost 20	MED = 20	MED = 0
redistribute isis route-map set metric 50	MED: IGP cost 50	MED: 50 MED: 50	MED: 50 MED: 50
redistribute isis metric 100	MED: IGP cost 100	MED: 100	MED: 100

Ignore Router-ID for Some Best-Path Calculations

Dell Networking OS allows you to avoid unnecessary BGP best-path transitions between external paths under certain conditions. The `bgp bestpath router-id ignore` command reduces network disruption caused by routing and forwarding plane changes and allows for faster convergence.

Four-Byte AS Numbers

Dell Networking OS supports 4-Byte (32-bit) format when configuring autonomous system numbers (ASNs).

The 4-Byte support is advertised as a new BGP capability (4-BYTE-AS) in the OPEN message. If a 4-Byte BGP speaker has sent and received this capability from another speaker, all the messages will be 4-octet. The behavior of a 4-Byte BGP speaker is different with the peer depending on whether the peer is a 4-Byte or 2-Byte BGP speaker.

Where the 2-Byte format is 1-65535, the 4-Byte format is 1-4294967295. Enter AS numbers using the traditional format. If the ASN is greater than 65535, the dot format is shown when using the `show ip bgp` commands. For example, an ASN entered as 3183856184 appears in the `show` commands as 48581.51768; an ASN of 65123 is shown as 65123. To calculate the comparable dot format for an ASN from a traditional format, use ASN/65536. ASN%65536.

Traditional Format	DOT Format
65001	0.65501
65536	1.0
100000	1.34464
4294967295	65535.65535

When creating Confederations, all the routers in a Confederation must be either 4-Byte or 2-Byte identified routers. You cannot mix them.

Configure 4-byte AS numbers with the `four-octet-support` command.

AS4 Number Representation

Dell Networking OS supports multiple representations of 4-byte AS numbers: asplain, asdot+, and asdot.

NOTE: The ASDOT and ASDOT+ representations are supported only with the 4-Byte AS numbers feature. If 4-Byte AS numbers are not implemented, only ASPLAIN representation is supported.

ASPLAIN is the method Dell Networking OS has used for all previous Dell Networking OS versions. ASPLAIN remains the default method with Dell Networking OS. With the ASPLAIN notation, a 32-bit binary AS number is translated into a decimal value.

- All AS numbers between 0 and 65535 are represented as a decimal number when entered in the CLI and when displayed in the `show` commands output.
- AS numbers larger than 65535 are represented using ASPLAIN notation. When entered in the CLI and when displayed in the `show` commands output, 65546 is represented as 65546.

ASDOT+ representation splits the full binary 4-byte AS number into two words of 16 bits separated by a decimal point (.): <high-order 16 bit value>.<low-order 16 bit value>. Some examples are shown in the following table.

- All AS numbers between 0 and 65535 are represented as a decimal number, when entered in the CLI and when displayed in the `show` commands outputs.
- AS Numbers larger than 65535 is represented using ASDOT notation as <higher 2 bytes in decimal>.<lower 2 bytes in decimal>. For example: AS 65546 is represented as 1.10.

ASDOT representation combines the ASPLAIN and ASDOT+ representations. AS numbers less than 65536 appear in integer format (asplain); AS numbers equal to or greater than 65536 appear in the decimal format (asdot+). For example, the AS number 65526 appears as 65526 and the AS number 65546 appears as 1.10.

Dynamic AS Number Notation Application

Dell Networking OS applies the ASN notation type change dynamically to the running-config statements.

When you apply or change an asnotation, the type selected is reflected immediately in the running-configuration and the `show` commands (refer to the following two examples).

Example of Dynamic Changes in the Running Configuration When Using the `bgp asnotation` Command

```
ASDOT
Dell(conf-router_bgp)#bgp asnotation asdot
Dell(conf-router_bgp)#show conf
!
router bgp 100
bgp asnotation asdot
bgp four-octet-as-support
neighbor 172.30.1.250 local-as 65057
<output truncated>

Dell(conf-router_bgp)#do show ip bgp
BGP table version is 24901, local router ID is 172.30.1.57
<output truncated>

ASDOT+
Dell(conf-router_bgp)#bgp asnotation asdot+
Dell(conf-router_bgp)#show conf
```

```
!
router bgp 100
bgp asnotation asdot+
bgp four-octet-as-support
neighbor 172.30.1.250 local-as 65057
<output truncated>

Dell(conf-router_bgp)#do show ip bgp
BGP table version is 31571, local router ID is 172.30.1.57
<output truncated>

AS-PLAIN
Dell(conf-router_bgp)#bgp asnotation asplain
Dell(conf-router_bgp)#sho conf
!
router bgp 100
bgp four-octet-as-support
neighbor 172.30.1.250 local-as 65057
<output truncated>

Dell(conf-router_bgp)#do sho ip bgp
BGP table version is 34558, local router ID is 172.30.1.57
<output truncated>
```

Example of the Running Configuration When AS Notation is Disabled

```
AS NOTATION DISABLED
Dell(conf-router_bgp)#no bgp asnotation
Dell(conf-router_bgp)#sho conf
!
router bgp 100
bgp four-octet-as-support
neighbor 172.30.1.250 local-as 65057
<output truncated>

Dell(conf-router_bgp)#do sho ip bgp
BGP table version is 28093, local router ID is 172.30.1.57
AS4 SUPPORT DISABLED
Dell(conf-router_bgp)#no bgp four-octet-as-support
Dell(conf-router_bgp)#sho conf
!
router bgp 100
neighbor 172.30.1.250 local-as 65057
Dell(conf-router_bgp)#do show ip bgp
BGP table version is 28093, local router ID is 172.30.1.57
```

AS Number Migration

With this feature you can transparently change the AS number of an entire BGP network and ensure that the routes are propagated throughout the network while the migration is in progress.

When migrating one AS to another, perhaps combining ASs, an eBGP network may lose its routing to an iBGP if the ASN changes. Migration can be difficult as all the iBGP and eBGP peers of the migrating network must be updated to maintain network reachability. Essentially, Local-AS provides a capability to the BGP speaker to operate as if it belongs to "virtual" AS network besides its physical AS network.

The following illustration shows a scenario where Router A, Router B, and Router C belong to AS 100, 200, and 300, respectively. Router A acquired Router B; Router B has Router C as its customer. When Router B is migrating to Router A, it must maintain the connection with Router C without immediately updating Router C’s configuration. Local-AS allows this behavior to happen by allowing Router B to appear as if it still belongs to Router B's old network (AS 200) as far as communicating with Router C is concerned.

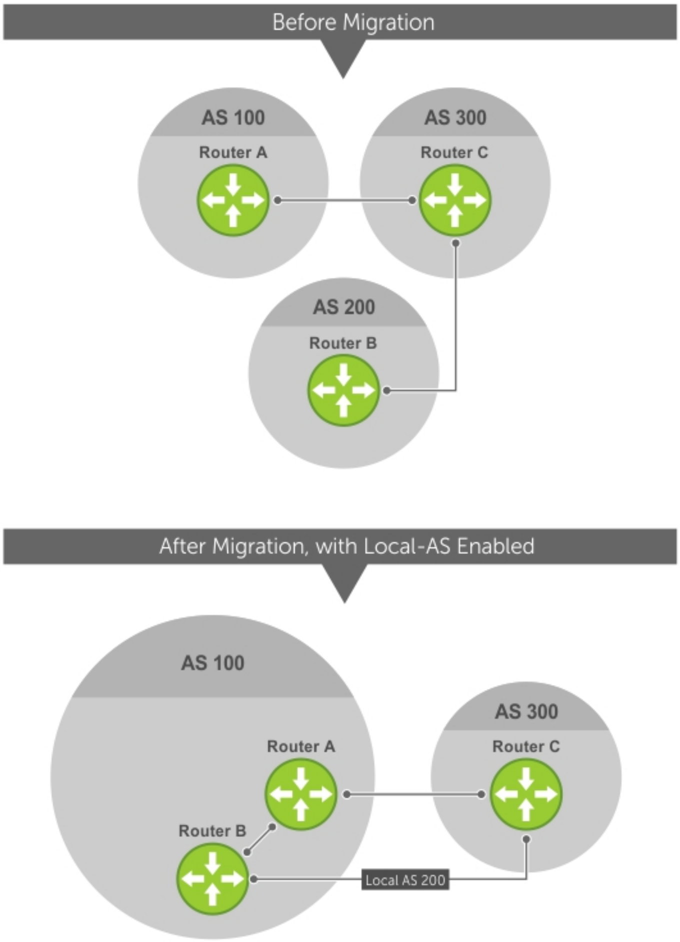

Before Migration
AS 100
Router A
AS 300
Router C
AS 200
Router B
After Migration, with Local-AS Enabled
AS 100
Router A
Router B
AS 300
Router C
Local AS 200

Figure 23. Before and After AS Number Migration with Local-AS Enabled

When you complete your migration, and you have reconfigured your network with the new information, disable this feature.

If you use the "no prepend" option, the Local-AS does not prepend to the updates received from the eBGP peer. If you do not select "no prepend" (the default), the Local-AS is added to the first AS segment in the AS-PATH. If an inbound route-map is used to prepend the as-path to the update from the peer, the Local-AS is added first. For example, consider the topology described in the previous illustration. If Router B has an inbound route-map applied on Router C to prepend "65001 65002" to the as-path, the following events take place on Router B:

1. Receive and validate the update.
2. Prepend local-as 200 to as-path.

3. Prepend "65001 65002" to as-path.

Local-AS is prepended before the route-map to give an impression that update passed through a router in AS 200 before it reached Router B.

BGP4 Management Information Base (MIB)

The FORCE10-BGP4-V2-MIB enhances Dell Networking OS BGP management information base (MIB) support with many new simple network management protocol (SNMP) objects and notifications (traps) defined in *draft-ietf-idr-bgp4-mibv2-05*. To see these enhancements, download the MIB from the Dell website.

NOTE: For the *Force10-BGP4-V2-MIB* and other MIB documentation, refer to the Dell iSupport web page.

Important Points to Remember

- Because eBGP packets are not controlled by the ACL, packets from BGP neighbors cannot be blocked using the `deny ip` command.
- The *f10BgpM2AsPathTableEntry* table, *f10BgpM2AsPathSegmentIndex*, and *f10BgpM2AsPathElementIndex* are used to retrieve a particular ASN from the AS path. These indices are assigned to the AS segments and individual ASN in each segment starting from 0. For example, an AS path list of {200 300 400} 500 consists of two segments: {200 300 400} with segment index 0 and 500 with segment index 1. ASN 200, 300, and 400 are assigned 0, 1, and 2 element indices in that order.
- Unknown optional transitive attributes within a given path attribute (PA) are assigned indices in order. These indices correspond to the *f10BgpM2PathAttrUnknownIndex* field in the *f10BgpM2PathAttrUnknownEntry* table.
- Negotiation of multiple instances of the same capability is not supported. *F10BgpM2PeerCapAnnouncedIndex* and *f10BgpM2PeerCapReceivedIndex* are ignored in the peer capability lookup.
- Configure inbound BGP soft-reconfiguration on a peer for *f10BgpM2PrefixInPrefixesRejected* to display the number of prefixes filtered due to a policy. If you do enable `BGP soft-reconfig`, the denied prefixes are not accounted for.
- *F10BgpM2AdjRibsOutRoute* stores the pointer to the NLRI in the peer's Adj-Rib-Out.
- PA Index (*f10BgpM2PathAttrIndex* field in various tables) is used to retrieve specific attributes from the PA table. The Next-Hop, RR Cluster-list, and Originator ID attributes are not stored in the PA Table and cannot be retrieved using the `index passed in` command. These fields are not populated in *f10BgpM2PathAttrEntry*, *f10BgpM2PathAttrClusterEntry*, and *f10BgpM2PathAttrOriginatorIdEntry*.
- *F10BgpM2PathAttrUnknownEntry* contains the optional-transitive attribute details.
- Query for *f10BgpM2LinkLocalNextHopEntry* returns the default value for Link-local Next-hop.
- RFC 2545 and the *f10BgpM2Rfc2545Group* are not supported.
- An SNMP query displays up to 89 AS paths. A query for a larger AS path count displays as "…" at the end of the output.
- SNMP set for BGP is not supported. For all peer configuration tables (*f10BgpM2PeerConfigurationGroup*, *f10BgpM2PeerRouteReflectorCfgGroup*, and *f10BgpM2PeerAsConfederationCfgGroup*), an SNMP set operation returns an error. Only SNMP queries are supported. In addition, the *f10BgpM2CfgPeerError*, *f10BgpM2CfgPeerBgpPeerEntry*, and *f10BgpM2CfgPeerRowEntryStatus* fields are to hold the SNMP set status and are ignored in SNMP query.
- The AFI/SAFI is not used as an index to the *f10BgpM2PeerCountersEntry* table. The BGP peer’s AFI/SAFI (IPv4 Unicast or IPv6 Multicast) is used for various outbound counters. Counters corresponding to IPv4 Multicast cannot be queried.

- The *f10BgpM2[Cfg]PeerReflectorClient* field is populated based on the assumption that route-reflector clients are not in a full mesh if you enable `BGP client-2-client reflection` and that the BGP speaker acting as reflector advertises routes learned from one client to another client. If disabled, it is assumed that clients are in a full mesh and there is no need to advertise prefixes to the other clients.
- High CPU utilization may be observed during an SNMP walk of a large BGP Loc-RIB.
- To avoid SNMP timeouts with a large-scale configuration (large number of BGP neighbors and a large BGP Loc-RIB), Dell Networking recommends setting the timeout and retry count values to a relatively higher number. For example, t = 60 or r = 5.
- To return all values on an snmpwalk for the *f10BgpM2Peer sub-OID*, use the `-C c` option, such as `snmpwalk -v 2c -C c -c public<IP_address><OID>`.
- An SNMP walk may terminate pre-maturely if the index does not increment lexicographically. Dell Networking recommends using options to ignore such errors.
- Multiple BPG process instances are not supported. Thus, the *f10BgpM2PeerInstance* field in various tables is not used to locate a peer.
- Multiple instances of the same NLRI in the BGP RIB are not supported and are set to zero in the SNMP query response.
- The *f10BgpM2NlriIndex* and *f10BgpM2AdjRibsOutIndex* fields are not used.
- Carrying MPLS labels in BGP is not supported. The *f10BgpM2NlriOpaqueType* and *f10BgpM2NlriOpaquePointer* fields are set to zero.
- 4-byte ASN is supported. The *f10BgpM2AsPath4byteEntry* table contains 4-byte ASN-related parameters based on the configuration.
- If a received update route matches with a local prefix, then that route is discarded. This behavior results from an incorrect BGP configuration. To overcome this issue, you can trigger a route refresh after you properly configure BGP.

Traps (notifications) specified in the BGP4 MIB draft `<draft-ietf-idr-bgp4-mibv2-05.txt>` are not supported. Such traps (*bgpM2Established* and *bgpM2BackwardTransition*) are supported as part of RFC 1657.

Configuration Information

The software supports BGPv4 as well as the following:

- deterministic multi-exit discriminator (MED) (default)
- a path with a missing MED is treated as worst path and assigned an MED value of (0xffffffff)
- the community format follows RFC 1998
- delayed configuration (the software at system boot reads the entire configuration file prior to sending messages to start BGP peer sessions)

The following are not yet supported:

- auto-summarization (the default is no auto-summary)
- synchronization (the default is no synchronization)

BGP Configuration

To enable the BGP process and begin exchanging information, assign an AS number and use commands in ROUTER BGP mode to configure a BGP neighbor.

By default, BGP is disabled.

By default, Dell Networking OS compares the MED attribute on different paths from within the same AS (the `bgp always-compare-med` command is not enabled).

NOTE: In Dell Networking OS, all newly configured neighbors and peer groups are disabled. To enable a neighbor or peer group, enter the `neighbor {ip-address | peer-group-name} no shutdown` command.

The following table displays the default values for BGP on Dell Networking OS.

Table 7. BGP Default Values

Item	Default
BGP Neighbor Adjacency changes	All BGP neighbor changes are logged.
Fast External Fallover feature	Disabled
Graceful Restart feature	Disabled
Local preference	100
MED	0
Route Flap Damping Parameters	half-life = 15 minutes reuse = 750 suppress = 2000 max-suppress-time = 60 minutes
Distance	external distance = 20 internal distance = 200 local distance = 200
Timers	keepalive = 60 seconds holdtime = 180 seconds
Add-path	Disabled

Enabling BGP

By default, BGP is not enabled on the system. Dell Networking OS supports one autonomous system (AS) and assigns the AS number (ASN).

To establish BGP sessions and route traffic, configure at least one BGP neighbor or peer.

In BGP, routers with an established TCP connection are called neighbors or peers. After a connection is established, the neighbors exchange full BGP routing tables with incremental updates afterward. In addition, neighbors exchange KEEPALIVE messages to maintain the connection.

In BGP, neighbor routers or peers can be classified as internal or external. External BGP peers must be connected physically to one another (unless you enable the EBGP multihop feature), while internal BGP peers do not need to be directly connected. The IP address of an EBGP neighbor is usually the IP address of the interface directly connected to the router. First, the BGP process determines if all internal BGP peers are reachable, then it determines which peers outside the AS are reachable.

NOTE: Sample Configurations for enabling BGP routers are found at the end of this chapter.

1. Assign an AS number and enter ROUTER BGP mode.

   CONFIGURATION mode

   ```
   router bgp as-number
   ```

   - *`as-number`*: from 0 to 65535 (2 Byte) or from 1 to 4294967295 (4 Byte) or 0.1 to 65535.65535 (Dotted format).

   Only one AS is supported per system.

   NOTE: If you enter a 4-Byte AS number, 4-Byte AS support is enabled automatically.

   a. Enable 4-Byte support for the BGP process.

      NOTE: This command is OPTIONAL. Enable if you want to use 4-Byte AS numbers or if you support AS4 number representation.

      CONFIG-ROUTER-BGP mode

      ```
      bgp four-octet-as-support
      ```

      NOTE: Use it only if you support 4-Byte AS numbers or if you support AS4 number representation. If you are supporting 4-Byte ASNs, enable this command.

      Disable 4-Byte support and return to the default 2-Byte format by using the `no bgp four-octet-as-support` command. You cannot disable 4-Byte support if you currently have a 4-Byte ASN configured.

      Disabling 4-Byte AS numbers also disables ASDOT and ASDOT+ number representation. All AS numbers are displayed in ASPLAIN format.

   b. Enable IPv4 multicast or IPv6 mode.

      CONFIG-ROUTER-BGP mode

      ```
      address-family [ipv4 | ipv6}
      ```

      Use this command to enter BGP for IPv6 mode (CONF-ROUTER_BGPv6_AF).

2. Add a neighbor as a remote AS.

   CONFIG-ROUTER-BGP mode

   ```
   neighbor {ip-address | peer-group name} remote-as as-number
   ```

   - *`peer-group name`*: 16 characters
   - *`as-number`*: from 0 to 65535 (2 Byte) or from 1 to 4294967295 (4 Byte) or 0.1 to 65535.65535 (Dotted format)

   Formats: IP Address A.B.C.D

   You must Configure Peer Groups *before* assigning it a remote AS.

3. Enable the BGP neighbor.

CONFIG-ROUTER-BGP mode

```
neighbor {ip-address | peer-group-name} no shutdown
```

Examples of the `show ip bgp` Commands

NOTE: When you change the configuration of a BGP neighbor, always reset it by entering the `clear ip bgp *` command in EXEC Privilege mode.

To view the BGP configuration, enter `show config` in CONFIGURATION ROUTER BGP mode. To view the BGP status, use the `show ip bgp summary` command in EXEC Privilege mode. The first example shows the summary with a 2-byte AS number displayed (in bold); the second example shows that the summary with a 4-byte AS number using the `show ip bgp summary` command (displays a 4–byte AS number in bold).

The following example shows the `show ip bgp summary` command output (2–byte AS number displays).

```
R2#show ip bgp summary
BGP router identifier 192.168.10.2, local AS number 65123
BGP table version is 1, main routing table version 1
1 network entrie(s) using 132 bytes of memory
1 paths using 72 bytes of memory
BGP-RIB over all using 73 bytes of memory
1 BGP path attribute entrie(s) using 72 bytes of memory
1 BGP AS-PATH entrie(s) using 47 bytes of memory
5 neighbor(s) using 23520 bytes of memory

Neighbor     AS    MsgRcvd MsgSent TblVer InQ OutQ Up/Down State/Pfx

10.10.21.1   65123 0       0       0      0    0 never    Active
10.10.32.3   65123 0       0       0      0    0 never    Active
100.10.92.9  65192 0       0       0      0    0 never    Active
192.168.10.1 65123 0       0       0      0    0 never    Active
192.168.12.2 65123 0       0       0      0    0 never    Active
R2#
```

The following example shows the `show ip bgp summary` command output (4–byte AS number displays).

```
R2#show ip bgp summary
BGP router identifier 192.168.10.2, local AS number 48735.59224
BGP table version is 1, main routing table version 1
1 network entrie(s) using 132 bytes of memory
1 paths using 72 bytes of memory
BGP-RIB over all using 73 bytes of memory
1 BGP path attribute entrie(s) using 72 bytes of memory
1 BGP AS-PATH entrie(s) using 47 bytes of memory
5 neighbor(s) using 23520 bytes of memory

Neighbor     AS    MsgRcvd MsgSent TblVer InQ OutQ Up/Down State/Pfx

10.10.21.1   65123 0       0       0      0   0    never   Active
10.10.32.3   65123 0       0       0      0   0    never   Active
100.10.92.9  65192 0       0       0      0   0    never   Active
192.168.10.1 65123 0       0       0      0   0    never   Active
192.168.12.2 65123 0       0       0      0   0    never   Active
R2#
```

For the router's identifier, Dell Networking OS uses the highest IP address of the Loopback interfaces configured. Because Loopback interfaces are virtual, they cannot go down, thus preventing changes in the router ID. If you do not configure Loopback interfaces, the highest IP address of any interface is used as the router ID.

To view the status of BGP neighbors, use the `show ip bgp neighbors` command in EXEC Privilege mode as shown in the first example. For BGP neighbor configuration information, use the `show running-config bgp` command in EXEC Privilege mode as shown in the second example.

NOTE: The `showconfig` command in CONFIGURATION ROUTER BGP mode gives the same information as the `show running-config bgp` command.

The following example displays two neighbors: one is an external internal BGP neighbor and the second one is an internal BGP neighbor. The first line of the output for each neighbor displays the AS number and states whether the link is an external or internal (shown in bold).

The third line of the `show ip bgp neighbors` output contains the BGP State. If anything other than ESTABLISHED is listed, the neighbor is not exchanging information and routes. For more information about using the `show ip bgp neighbors` command, refer to the *Dell Networking OS Command Line Interface Reference Guide*.

The following example shows the `show ip bgp neighbors` command output.

```
Dell#show ip bgp neighbors
```

BGP neighbor is 10.114.8.60, remote AS 18508, external link

```
  BGP version 4, remote router ID 10.20.20.20
  BGP state ESTABLISHED, in this state for 00:01:58
  Last read 00:00:14, hold time is 90, keepalive interval is 30 seconds
  Received 18552 messages, 0 notifications, 0 in queue
  Sent 11568 messages, 0 notifications, 0 in queue
  Received 18549 updates, Sent 11562 updates
  Minimum time between advertisement runs is 30 seconds

  For address family: IPv4 Unicast
  BGP table version 216613, neighbor version 201190
  130195 accepted prefixes consume 520780 bytes
  Prefix advertised 49304, rejected 0, withdrawn 36143

  Connections established 1; dropped 0
  Last reset never
Local host: 10.114.8.39, Local port: 1037
Foreign host: 10.114.8.60, Foreign port: 179
```

BGP neighbor is 10.1.1.1, remote AS 65535, internal link

```
  Administratively shut down
  BGP version 4, remote router ID 10.0.0.0
  BGP state IDLE, in this state for 17:12:40
  Last read 17:12:40, hold time is 180, keepalive interval is 60 seconds
  Received 0 messages, 0 notifications, 0 in queue
  Sent 0 messages, 0 notifications, 0 in queue
  Received 0 updates, Sent 0 updates
  Minimum time between advertisement runs is 5 seconds

  For address family: IPv4 Unicast
  BGP table version 0, neighbor version 0
  0 accepted prefixes consume 0 bytes
  Prefix advertised 0, rejected 0, withdrawn 0
  Connections established 0; dropped 0
  Last reset never
  No active TCP connection
Dell#
```

The following example shows verifying the BGP configuration using the `show running-config bgp` command..

```
Dell#show running-config bgp
!
router bgp 65123
  bgp router-id 192.168.10.2
  network 10.10.21.0/24
  network 10.10.32.0/24
  network 100.10.92.0/24
  network 192.168.10.0/24
  bgp four-octet-as-support
  neighbor 10.10.21.1 remote-as 65123
  neighbor 10.10.21.1 filter-list ISP1in
  neighbor 10.10.21.1 no shutdown
  neighbor 10.10.32.3 remote-as 65123
  neighbor 10.10.32.3 no shutdown
  neighbor 100.10.92.9 remote-as 65192
  neighbor 100.10.92.9 no shutdown
  neighbor 192.168.10.1 remote-as 65123
  neighbor 192.168.10.1 update-source Loopback 0
  neighbor 192.168.10.1 no shutdown
  neighbor 192.168.12.2 remote-as 65123
  neighbor 192.168.12.2 update-source Loopback 0
  neighbor 192.168.12.2 no shutdown
Dell#
```

Configuring AS4 Number Representations

Enable one type of AS number representation: ASPLAIN, ASDOT+, or ASDOT.

Term	Description
ASPLAIN	the method Dell Networking OS used for all previous Dell Networking OS versions. It remains the default method with Dell Networking OS. With the ASPLAIN notation, a 32–bit binary AS number is translated into a decimal value.
ASDOT+	representation splits the full binary 4-byte AS number into two words of 16 bits separated by a decimal point (.): <high-order 16 bit value>.<low-order 16 bit value>.
ASDOT	representation combines the ASPLAIN and ASDOT+ representations. AS numbers less than 65536 appear in integer format (asplain); AS numbers equal to or greater than 65536 appear using the decimal method (asdot+). For example, the AS number 65526 appears as 65526 and the AS number 65546 appears as 1.10.

NOTE: The ASDOT and ASDOT+ representations are supported only with the 4-Byte AS numbers feature. If you do not implement 4-Byte AS numbers, only ASPLAIN representation is supported.

Only one form of AS number representation is supported at a time. You cannot combine the types of representations within an AS.

To configure AS4 number representations, use the following commands.

- Enable ASPLAIN AS Number representation.

  CONFIG-ROUTER-BGP mode

```
bgp asnotation asplain
```

NOTE: ASPLAIN is the default method Dell Networking OS uses and does not appear in the configuration display.

- Enable ASDOT AS Number representation.

  CONFIG-ROUTER-BGP mode

  ```
  bgp asnotation asdot
  ```
- Enable ASDOT+ AS Number representation.

  CONFIG-ROUTER-BGP mode

  ```
  bgp asnotation asdot+
  ```

Examples of the `bgp asnotation` Commands

The following example shows the `bgp asnotation asplain` command output.

```
Dell(conf-router_bgp)#bgp asnotation asplain
Dell(conf-router_bgp)#sho conf
!
router bgp 100
  bgp four-octet-as-support
  neighbor 172.30.1.250 remote-as 18508
  neighbor 172.30.1.250 local-as 65057
  neighbor 172.30.1.250 route-map rmap1 in
  neighbor 172.30.1.250 password 7
5ab3eb9a15ed02ff4f0dfd4500d6017873cfd9a267c04957
  neighbor 172.30.1.250 no shutdown
5332332 9911991 65057 18508 12182 7018 46164 i
```

The following example shows the `bgp asnotation asdot` command output.

```
Dell(conf-router_bgp)#bgp asnotation asdot
Dell(conf-router_bgp)#sho conf
!
router bgp 100
  bgp asnotation asdot
bgp four-octet-as-support
  neighbor 172.30.1.250 remote-as 18508
  neighbor 172.30.1.250 local-as 65057
  neighbor 172.30.1.250 route-map rmap1 in
  neighbor 172.30.1.250 password 7
5ab3eb9a15ed02ff4f0dfd4500d6017873cfd9a267c04957
  neighbor 172.30.1.250 no shutdown
5332332 9911991 65057 18508 12182 7018 46164 i
```

The following example shows the `bgp asnotation asdot+` command output.

```
Dell(conf-router_bgp)#bgp asnotation asdot+
Dell(conf-router_bgp)#sho conf
!
router bgp 100
  bgp asnotation asdot+
bgp four-octet-as-support
  neighbor 172.30.1.250 remote-as 18508
  neighbor 172.30.1.250 local-as 65057
  neighbor 172.30.1.250 route-map rmap1 in
  neighbor 172.30.1.250 password 7
5ab3eb9a15ed02ff4f0dfd4500d6017873cfd9a267c04957
  neighbor 172.30.1.250 no shutdown
5332332 9911991 65057 18508 12182 7018 46164 i
```

Configuring Peer Groups

To configure multiple BGP neighbors at one time, create and populate a BGP peer group.
An advantage of peer groups is that members of a peer group inherit the configuration properties of the group and share same update policy.

A maximum of 256 peer groups are allowed on the system.

Create a peer group by assigning it a name, then adding members to the peer group. After you create a peer group, you can configure route policies for it. For information about configuring route policies for a peer group, refer to Filtering BGP Routes.

NOTE: Sample Configurations for enabling peer groups are found at the end of this chapter.

1. Create a peer group by assigning a name to it.

   CONFIG-ROUTERBGP mode

   ```
   neighbor peer-group-name peer-group
   ```
2. Enable the peer group.

   CONFIG-ROUTERBGP mode

   ```
   neighbor peer-group-name no shutdown
   ```

   By default, all peer groups are disabled.
3. Create a BGP neighbor.

   CONFIG-ROUTERBGP mode

   ```
   neighbor ip-address remote-as as-number
   ```
4. Enable the neighbor.

   CONFIG-ROUTERBGP mode

   ```
   neighbor ip-address no shutdown
   ```
5. Add an enabled neighbor to the peer group.

   CONFIG-ROUTERBGP mode

   ```
   neighbor ip-address peer-group peer-group-name
   ```
6. Add a neighbor as a remote AS.

   CONFIG-ROUTERBGP mode

   ```
   neighbor {ip-address | peer-group name} remote-as as-number
   ```

   Formats: IP Address A.B.C.D

   - `Peer-Group Name`: 16 characters.
   - `as-number`: the range is from 0 to 65535 (2-Byte) or 1 to 4294967295 | 0.1 to 65535.65535 (4-Byte) or 0.1 to 65535.65535 (Dotted format)

   To add an external BGP (EBGP) neighbor, configure the `as-number` parameter with a number different from the BGP as-number configured in the `router bgp as-number` command.

To add an internal BGP (IBGP) neighbor, configure the `as-number` parameter with the same BGP as-number configured in the `router bgp as-number` command.

Examples of Viewing and Configuring Peer Groups

After you create a peer group, you can use any of the commands beginning with the keyword `neighbor` to configure that peer group.

When you add a peer to a peer group, it inherits all the peer group's configured parameters.

A neighbor cannot become part of a peer group if it has any of the following commands configured:

- `neighbor advertisement-interval`
- `neighbor distribute-list out`
- `neighbor filter-list out`
- `neighbor next-hop-self`
- `neighbor route-map out`
- `neighbor route-reflector-client`
- `neighbor send-community`

A neighbor may keep its configuration after it was added to a peer group if the neighbor's configuration is more specific than the peer group's and if the neighbor's configuration does not affect outgoing updates.

NOTE: When you configure a new set of BGP policies for a peer group, *always* reset the peer group by entering the `clear ip bgp peer-group` *`peer-group-name`* command in EXEC Privilege mode.

To view the configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode.

When you create a peer group, it is disabled (shutdown). The following example shows the creation of a peer group (zanzibar) (in bold).

```
Dell(conf-router_bgp)#neighbor zanzibar peer-group
Dell(conf-router_bgp)#show conf
!
router bgp 45
  bgp fast-external-fallover
  bgp log-neighbor-changes
  neighbor zanzibar peer-group
  neighbor zanzibar shutdown
  neighbor 10.1.1.1 remote-as 65535
  neighbor 10.1.1.1 shutdown
  neighbor 10.14.8.60 remote-as 18505
  neighbor 10.14.8.60 no shutdown
Dell(conf-router_bgp)#
```

To enable a peer group, use the `neighbor` *`peer-group-name`* `no shutdown` command in CONFIGURATION ROUTER BGP mode (shown in bold).

```
Dell(conf-router_bgp)#neighbor zanzibar no shutdown
Dell(conf-router_bgp)#show config
!
router bgp 45
  bgp fast-external-fallover
  bgp log-neighbor-changes
  neighbor zanzibar peer-group
  neighbor zanzibar no shutdown
  neighbor 10.1.1.1 remote-as 65535
  neighbor 10.1.1.1 shutdown
  neighbor 10.14.8.60 remote-as 18505
  neighbor 10.14.8.60 no shutdown
Dell(conf-router_bgp)#
```

To disable a peer group, use the `neighbor` *`peer-group-name`* `shutdown` command in CONFIGURATION ROUTER BGP mode. The configuration of the peer group is maintained, but it is not applied to the peer group members. When you disable a peer group, all the peers within the peer group that are in the ESTABLISHED state move to the IDLE state.

To view the status of peer groups, use the `show ip bgp peer-group` command in EXEC Privilege mode, as shown in the following example.

```
Dell>show ip bgp peer-group

Peer-group zanzibar, remote AS 65535
BGP version 4
Minimum time between advertisement runs is 5 seconds

For address family: IPv4 Unicast
BGP neighbor is zanzibar, peer-group internal,
Number of peers in this group 26
Peer-group members (* - outbound optimized):
  10.68.160.1
  10.68.161.1
  10.68.162.1
  10.68.163.1
  10.68.164.1
  10.68.165.1
  10.68.166.1
  10.68.167.1
  10.68.168.1
  10.68.169.1
  10.68.170.1
  10.68.171.1
  10.68.172.1
  10.68.173.1
  10.68.174.1
  10.68.175.1
  10.68.176.1
  10.68.177.1
  10.68.178.1
  10.68.179.1
  10.68.180.1
  10.68.181.1
  10.68.182.1
  10.68.183.1
  10.68.184.1
  10.68.185.1
Dell>
```

Configuring BGP Fast Fall-Over

By default, a BGP session is governed by the hold time.
BGP routers typically carry large routing tables, so frequent session resets are not desirable. The BGP fast fall-over feature reduces the convergence time while maintaining stability. The connection to a BGP peer is immediately reset if a link to a directly connected external peer fails.

When you enable fall-over, BGP tracks IP reachability to the peer remote address and the peer local address. Whenever either address becomes unreachable (for example, no active route exists in the routing table for peer IPv6 destinations/local address), BGP brings down the session with the peer.

The BGP fast fall-over feature is configured on a per-neighbor or peer-group basis and is disabled by default.

To enable the BGP fast fall-over feature, use the following command.

To disable fast fall-over, use the `[no] neighbor [neighbor | peer-group] fall-over` command in CONFIGURATION ROUTER BGP mode.

- Enable BGP Fast fall-Over.
  CONFIG-ROUTER-BGP mode

  `neighbor {`*`ip-address`*` | `*`peer-group-name`*`} fall-over`

Examples of Verifying that Fast fall-Over is Enabled on a BGP Neighbor and a Peer-Group

To verify that you enabled fast fall-over on a particular BGP neighbor, use the `show ip bgp neighbors` command. Because fast fall-over is disabled by default, it appears only if it has been enabled (shown in bold).

```
Dell#sh ip bgp neighbors

BGP neighbor is 100.100.100.100, remote AS 65517, internal link
  Member of peer-group test for session parameters
  BGP version 4, remote router ID 30.30.30.5
  BGP state ESTABLISHED, in this state for 00:19:15
  Last read 00:00:15, last write 00:00:06
  Hold time is 180, keepalive interval is 60 seconds
  Received 52 messages, 0 notifications, 0 in queue
  Sent 45 messages, 5 notifications, 0 in queue
  Received 6 updates, Sent 0 updates
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 5 seconds
  Minimum time before advertisements start is 0 seconds

  Capabilities received from neighbor for IPv4 Unicast :
    MULTIPROTO_EXT(1)
    ROUTE_REFRESH(2)
    CISCO_ROUTE_REFRESH(128)

  Capabilities advertised to neighbor for IPv4 Unicast :
    MULTIPROTO_EXT(1)
    ROUTE_REFRESH(2)
    CISCO_ROUTE_REFRESH(128)
```

fall-over enabled

```
Update source set to Loopback 0

Peer active in peer-group outbound optimization

For address family: IPv4 Unicast
BGP table version 52, neighbor version 52
4 accepted prefixes consume 16 bytes
Prefix advertised 0, denied 0, withdrawn 0

Connections established 6; dropped 5
Last reset 00:19:37, due to Reset by peer
```

```
Notification History
  'Connection Reset' Sent : 5 Recv: 0

Local host: 200.200.200.200, Local port: 65519
Foreign host: 100.100.100.100, Foreign port: 179

Dell#
```

To verify that fast fall-over is enabled on a peer-group, use the `show ip bgp peer-group` command (shown in bold).

```
Dell#sh ip bgp peer-group

Peer-group test
  fall-over enabled
  BGP version 4
  Minimum time between advertisement runs is 5 seconds

  For address family: IPv4 Unicast
  BGP neighbor is test
  Number of peers in this group 1
  Peer-group members (* - outbound optimized):
    100.100.100.100*

Dell#

router bgp 65517
  neighbor test peer-group
  neighbor test fall-over
  neighbor test no shutdown
  neighbor 100.100.100.100 remote-as 65517
  neighbor 100.100.100.100 fall-over
  neighbor 100.100.100.100 update-source Loopback 0
  neighbor 100.100.100.100 no shutdown
Dell#
```

Configuring Passive Peering

When you enable a peer-group, the software sends an OPEN message to initiate a TCP connection. If you enable passive peering for the peer group, the software does not send an OPEN message, but it responds to an OPEN message.

When a BGP neighbor connection with authentication configured is rejected by a passive peer-group, Dell Networking OS does not allow another passive peer-group on the same subnet to connect with the BGP neighbor. To work around this, change the BGP configuration or change the order of the peer group configuration.

You can constrain the number of passive sessions accepted by the neighbor. The `limit` keyword allows you to set the total number of sessions the neighbor will accept, between 2 and 265. The default is **256** sessions.

1. Configure a peer group that does not initiate TCP connections with other peers.

   CONFIG-ROUTER-BGP mode

   ```
   neighbor peer-group-name peer-group passive limit
   ```

   Enter the `limit` keyword to restrict the number of sessions accepted.
2. Assign a subnet to the peer group.

CONFIG-ROUTER-BGP mode

`neighbor peer-group-name subnet subnet-number mask`

The peer group responds to OPEN messages sent on this subnet.

3. Enable the peer group.

   CONFIG-ROUTER-BGP mode

   `neighbor peer-group-name no shutdown`
4. Create and specify a remote peer for BGP neighbor.

   CONFIG-ROUTER-BGP mode

   `neighbor peer-group-name remote-as as-number`

Only after the peer group responds to an OPEN message sent on the subnet does its BGP state change to ESTABLISHED. After the peer group is ESTABLISHED, the peer group is the same as any other peer group.

For more information about peer groups, refer to Configure Peer Groups.

Maintaining Existing AS Numbers During an AS Migration

The local-as feature smooths out the BGP network migration operation and allows you to maintain existing ASNs during a BGP network migration.
When you complete your migration, be sure to reconfigure your routers with the new information and disable this feature.

- Allow external routes from this neighbor.

  CONFIG-ROUTERBGP mode

  `neighbor {IP address | peer-group-name local-as as number [no prepend]`

  - `Peer Group Name`: 16 characters.
  - `AS-number`: 0 to 65535 (2-Byte) or 1 to 4294967295 (4-Byte) or 0.1 to 65535.65535 (Dotted format).
  - `No Prepend`: specifies that local AS values are not prepended to announcements from the neighbor.

  Format: IP Address: A.B.C.D.

  You must Configure Peer Groups *before* assigning it to an AS. This feature is not supported on passive peer groups.

Example of the Verifying that Local AS Numbering is Disabled

The first line in bold shows the actual AS number. The second two lines in bold show the local AS number (6500) maintained during migration.

To disable this feature, use the `no neighbor local-as` command in CONFIGURATION ROUTER BGP mode.

```
R2(conf-router_bgp)#show conf
!
router bgp 65123
  bgp router-id 192.168.10.2
  network 10.10.21.0/24
  network 10.10.32.0/24
  network 100.10.92.0/24
  network 192.168.10.0/24
  bgp four-octet-as-support
  neighbor 10.10.21.1 remote-as 65123
  neighbor 10.10.21.1 filter-list Laura in
  neighbor 10.10.21.1 no shutdown
  neighbor 10.10.32.3 remote-as 65123
  neighbor 10.10.32.3 no shutdown
  neighbor 100.10.92.9 remote-as 65192
 neighbor 100.10.92.9 local-as 6500
 neighbor 100.10.92.9 no shutdown
  neighbor 192.168.10.1 remote-as 65123
  neighbor 192.168.10.1 update-source Loopback 0
  neighbor 192.168.10.1 no shutdown
  neighbor 192.168.12.2 remote-as 65123
  neighbor 192.168.12.2 update-source Loopback 0
  neighbor 192.168.12.2 no shutdown
R2(conf-router_bgp)#
```

Allowing an AS Number to Appear in its Own AS Path

This command allows you to set the number of times a particular AS number can occur in the AS path. The allow-as feature permits a BGP speaker to allow the ASN to be present for a specified number of times in the update received from the peer, even if that ASN matches its own. The AS-PATH loop is detected if the local ASN is present more than the specified number of times in the command.

- Allow this neighbor ID to use the AS path the specified number of times.

  CONFIG-ROUTER-BGP mode

  ```
  neighbor {IP address | peer-group-name} allowas-in number
  ```

  - *Peer Group Name*: 16 characters.
  - *Number*: 1 through 10.

  Format: IP Address: A.B.C.D.

  You must Configure Peer Groups before assigning it to an AS.

Example of Viewing AS Numbers in AS Paths

The lines shown in bold are the number of times ASN 65123 can appear in the AS path (**allows–in 9**).

To disable this feature, use the `no neighbor allow-as in number` command in CONFIGURATION ROUTER BGP mode.

```
R2(conf-router_bgp)#show conf
!
router bgp 65123
  bgp router-id 192.168.10.2
  network 10.10.21.0/24
  network 10.10.32.0/24
  network 100.10.92.0/24
  network 192.168.10.0/24
  bgp four-octet-as-support
  neighbor 10.10.21.1 remote-as 65123
  neighbor 10.10.21.1 filter-list Laura in
  neighbor 10.10.21.1 no shutdown
  neighbor 10.10.32.3 remote-as 65123
  neighbor 10.10.32.3 no shutdown
  neighbor 100.10.92.9 remote-as 65192
```

```
  neighbor 100.10.92.9 local-as 6500
  neighbor 100.10.92.9 no shutdown
  neighbor 192.168.10.1 remote-as 65123
  neighbor 192.168.10.1 update-source Loopback 0
  neighbor 192.168.10.1 no shutdown
  neighbor 192.168.12.2 remote-as 65123
 neighbor 192.168.12.2 allowas-in 9
 neighbor 192.168.12.2 update-source Loopback 0
 neighbor 192.168.12.2 no shutdown
R2(conf-router_bgp)#R2(conf-router_bgp)#
```

Enabling Graceful Restart

Use this feature to lessen the negative effects of a BGP restart.
Dell Networking OS advertises support for this feature to BGP neighbors through a capability advertisement. You can enable graceful restart by router and/or by peer or peer group.

NOTE: By default, BGP graceful restart is disabled.

The default role for BGP is as a receiving or restarting peer. If you enable BGP, when a peer that supports graceful restart resumes operating, Dell Networking OS performs the following tasks:

- Continues saving routes received from the peer if the peer advertised it had graceful restart capability. Continues forwarding traffic to the peer.
- Flags routes from the peer as Stale and sets a timer to delete them if the peer does not perform a graceful restart.
- Deletes all routes from the peer if forwarding state information is not saved.
- Speeds convergence by advertising a special update packet known as an end-of-RIB marker. This marker indicates the peer has been updated with all routes in the local RIB.

If you configure your system to do so, Dell Networking OS can perform the following actions during a hot failover:

- Save all forwarding information base (FIB) and content addressable memory (CAM) entries on the line card and continue forwarding traffic while the secondary route processor module (RPM) is coming online.
- Advertise to all BGP neighbors and peer-groups that the forwarding state of all routes has been saved. This prompts all peers to continue saving the routes they receive and to continue forwarding traffic.
- Bring the secondary RPM online as the primary and re-open sessions with all peers operating in No Shutdown mode.
- Defer best path selection for a certain amount of time. This helps optimize path selection and results in fewer updates being sent out.

To enable graceful restart, use the `configure router bgp graceful-restart` command.

- Enable graceful restart for the BGP node.

  CONFIG-ROUTER-BGP mode

  ```
  bgp graceful-restart
  ```
- Set maximum restart time for all peers.

  CONFIG-ROUTER-BGP mode

  ```
  bgp graceful-restart [restart-time time-in-seconds]
  ```

  The default is **120 seconds**.

- Set maximum time to retain the restarting peer's stale paths.

  CONFIG-ROUTER-BGP mode

  ```
  bgp graceful-restart [stale-path-time time-in-seconds]
  ```

  The default is **360 seconds**.
- Local router supports graceful restart as a receiver only.

  CONFIG-ROUTER-BGP mode

  ```
  bgp graceful-restart [role receiver-only]
  ```

Enabling Neighbor Graceful Restart

BGP graceful restart is active only when the neighbor becomes established. Otherwise, it is disabled. Graceful-restart applies to all neighbors with established adjacency.

With the graceful restart feature, Dell Networking OS enables the receiving/restarting mode by default. In Receiver-Only mode, graceful restart saves the advertised routes of peers that support this capability when they restart. This option provides support for remote peers for their graceful restart without supporting the feature itself.

You can implement BGP graceful restart either by neighbor or by BGP peer-group. For more information, refer to the *Dell Networking OS Command Line Interface Reference Guide*.

- Add graceful restart to a BGP neighbor or peer-group.

  CONFIG-ROUTER-BGP mode

  ```
  neighbor {ip-address | peer-group-name} graceful-restart
  ```
- Set the maximum restart time for the neighbor or peer-group.

  CONFIG-ROUTER-BGP mode

  ```
  neighbor {ip-address | peer-group-name} graceful-restart [restart-time time-
  in-seconds]
  ```

  The default is **120 seconds**.
- Local router supports graceful restart for this neighbor or peer-group as a receiver only.

  CONFIG-ROUTER-BGP mode

  ```
  neighbor {ip-address | peer-group-name} graceful-restart [role receiver-only]
  ```
- Set the maximum time to retain the restarting neighbor's or peer-group's stale paths.

  CONFIG-ROUTER-BGP mode

  ```
  neighbor {ip-address | peer-group-name} graceful-restart [stale-path-time
  time-in-seconds]
  ```

  The default is **360 seconds**.

Filtering on an AS-Path Attribute

You can use the BGP attribute, AS_PATH, to manipulate routing policies.

The AS_PATH attribute contains a sequence of AS numbers representing the route's path. As the route traverses an AS, the ASN is prepended to the route. You can manipulate routes based on their AS_PATH to affect interdomain routing. By identifying certain ASN in the AS_PATH, you can permit or deny routes based on the number in its AS_PATH.

AS-PATH ACLs use regular expressions to search AS_PATH values. AS-PATH ACLs have an “implicit deny.” This means that routes that do not meet a deny or match filter are dropped.

To configure an AS-PATH ACL to filter a specific AS_PATH value, use these commands in the following sequence.

1. Assign a name to a AS-PATH ACL and enter AS-PATH ACL mode.
   CONFIGURATION mode

   ```
   ip as-path access-list as-path-name
   ```
2. Enter the parameter to match BGP AS-PATH for filtering.
   CONFIG-AS-PATH mode

   ```
   {deny | permit} filter parameter
   ```

   This is the filter that is used to match the AS-path. The entries can be any format, letters, numbers, or regular expressions.

   You can enter this command multiple times if multiple filters are desired.

   For accepted expressions, refer to Regular Expressions as Filters.
3. Return to CONFIGURATION mode.
   AS-PATH ACL mode

   ```
   exit
   ```
4. Enter ROUTER BGP mode.
   CONFIGURATION mode

   ```
   router bgp as-number
   ```
5. Use a configured AS-PATH ACL for route filtering and manipulation.
   CONFIG-ROUTER-BGP mode

   ```
   neighbor {ip-address | peer-group-name} filter-list as-path-name {in | out}
   ```

   If you assign an non-existent or empty AS-PATH ACL, the software allows all routes.

Example of the `show ip bgp paths` Command

To view all BGP path attributes in the BGP database, use the `show ip bgp paths` command in EXEC Privilege mode.

```
Dell#show ip bgp paths
Total 30655 Paths
Address    Hash Refcount Metric Path
0x4014154 0    3        18508 701 3549 19421 i
0x4013914 0    3        18508 701 7018 14990 i
0x5166d6c 0    3        18508 209 4637 1221 9249 9249 i
0x5e62df4 0    2        18508 701 17302 i
0x3a1814c 0    26       18508 209 22291 i
0x567ea9c 0    75       18508 209 3356 2529 i
0x6cc1294 0    2        18508 209 1239 19265 i
```

```
0x6cc18d4 0    1        18508 701 2914 4713 17935 i
0x5982e44 0    162      18508 209 i
0x67d4a14 0    2        18508 701 19878 ?
0x559972c 0    31       18508 209 18756 i
0x59cd3b4 0    2        18508 209 7018 15227 i
0x7128114 0    10       18508 209 3356 13845 i
0x536a914 0    3        18508 209 701 6347 7781 i
0x2ffe884 0    1        18508 701 3561 9116 21350 i
0x2ff7284 0    99       18508 701 1239 577 855 ?
0x2ff7ec4 0    4        18508 209 3561 4755 17426 i
0x2ff8544 0    3        18508 701 5743 2648 i
0x736c144 0    1        18508 701 209 568 721 1494 i
0x3b8d224 0    10       18508 209 701 2019 i
0x5eb1e44 0    1        18508 701 8584 16158 i
0x5cd891c 0    9        18508 209 6453 4759 i
--More--
```

Regular Expressions as Filters

Regular expressions are used to filter AS paths or community lists. A regular expression is a special character used to define a pattern that is then compared with an input string.

For an AS-path access list, as shown in the previous commands, if the AS path matches the regular expression in the access list, the route matches the access list.

The following lists the regular expressions accepted in Dell Networking OS.

Regular Expression	Definition
^ (caret)	Matches the beginning of the input string. Alternatively, when used as the first character within brackets [^], this matches any number except the ones specified within the brackets.
$ (dollar)	Matches the end of the input string.
. (period)	Matches any single character, including white space.
* (asterisk)	Matches 0 or more sequences of the immediately previous character or pattern.
+ (plus)	Matches 1 or more sequences of the immediately previous character or pattern.
? (question)	Matches 0 or 1 sequence of the immediately previous character or pattern.
( ) (parenthesis)	Specifies patterns for multiple use when one of the multiplier metacharacters follows: asterisk *, plus sign +, or question mark ?
[ ] (brackets)	Matches any enclosed character and specifies a range of single characters.
- (hyphen)	Used within brackets to specify a range of AS or community numbers.
_ (underscore)	Matches a ^, a $, a comma, a space, or a {, or a }. Placed on either side of a string to specify a literal and disallow substring matching. You can precede or follow numerals enclosed by underscores by any of the characters listed.
| (pipe)	Matches characters on either side of the metacharacter; logical OR.

As seen in the following example, the expressions are displayed when using the `show` commands. To view the AS-PATH ACL configuration, use the `show config` command in CONFIGURATION AS-PATH ACL mode and the `show ip as-path-access-list` command in EXEC Privilege mode.

For more information about this command and route filtering, refer to Filtering BGP Routes.

The following example applies access list Eagle to routes inbound from BGP peer 10.5.5.2. Access list Eagle uses a regular expression to deny routes originating in AS 32. The first lines shown in bold create the access list and filter. The second lines shown in bold are the regular expression shown as part of the access list filter.

Example of Using Regular Expression to Filter AS Paths

```
Dell(config)#router bgp 99
Dell(conf-router_bgp)#neigh AAA peer-group
Dell(conf-router_bgp)#neigh AAA no shut
Dell(conf-router_bgp)#show conf
!
router bgp 99
  neighbor AAA peer-group
  neighbor AAA no shutdown
  neighbor 10.155.15.2 remote-as 32
  neighbor 10.155.15.2 shutdown
Dell(conf-router_bgp)#neigh 10.155.15.2 filter-list 1 in
Dell(conf-router_bgp)#ex

Dell(conf)#ip as-path access-list Eagle
Dell(config-as-path)#deny 32$
Dell(config-as-path)#ex
Dell(conf)#router bgp 99
Dell(conf-router_bgp)#neighbor AAA filter-list Eagle in
Dell(conf-router_bgp)#show conf
!
router bgp 99
  neighbor AAA peer-group
  neighbor AAA filter-list Eaglein
  neighbor AAA no shutdown
  neighbor 10.155.15.2 remote-as 32
  neighbor 10.155.15.2 filter-list 1 in
  neighbor 10.155.15.2 shutdown
Dell(conf-router_bgp)#ex
Dell(conf)#ex

Dell#show ip as-path-access-lists
ip as-path access-list Eagle
 deny 32$
Dell#
```

Redistributing Routes

In addition to filtering routes, you can add routes from other routing instances or protocols to the BGP process. With the `redistribute` command, you can include ISIS, OSPF, static, or directly connected routes in the BGP process.
To add routes from other routing instances or protocols, use any of the following commands in ROUTER BGP mode.

- Include, directly connected or user-configured (static) routes in BGP.

  ROUTER BGP or CONF-ROUTER_BGPv6_ AF mode

  ```
  redistribute {connected | static} [route-map map-name]
  ```

  Configure the `map-name` parameter to specify the name of a configured route map.
- Include specific ISIS routes in BGP.

  ROUTER BGP or CONF-ROUTER_BGPv6_ AF mode

```
redistribute isis [level-1 | level-1-2 | level-2] [metric value] [route-map
map-name]
```

Configure the following parameters:

- `level-1`, `level-1-2`, or `level-2`: Assign all redistributed routes to a level. The default is **level-2**.
- `metric value`: The value is from 0 to 16777215. The default is **0**.
- `map-name`: name of a configured route map.

- Include specific OSPF routes in IS-IS.
  ROUTER BGP or CONF-ROUTER_BGPv6_ AF mode

```
redistribute ospf process-id [match external {1 | 2} | match internal]
[metric-type {external | internal}] [route-map map-name]
```

Configure the following parameters:

- *`process-id`*: the range is from 1 to 65535.
- `match external`: the range is from 1 or 2.
- `match internal`
- `metric-type`: external or internal.
- `map-name`: name of a configured route map.

Enabling Additional Paths

The add-path feature is disabled by default.

NOTE: Dell Networking OS recommends *not* using multipath and add path simultaneously in a route reflector.

To allow multiple paths sent to peers, use the following commands.

1. Allow the advertisement of multiple paths for the same address prefix without the new paths replacing any previous ones.
   CONFIG-ROUTER-BGP mode

   ```
   bgp add-path [both|received|send] path-count count
   ```

   The range is from 2 to 64.
2. Allow the specified neighbor/peer group to send/ receive multiple path advertisements.
   CONFIG-ROUTER-BGP mode

   ```
   neighbor add-path
   ```

NOTE: The `path-count` parameter controls the number of paths that are advertised, not the number of paths that are received.

Configuring IP Community Lists

Within Dell Networking OS, you have multiple methods of manipulating routing attributes.
One attribute you can manipulate is the COMMUNITY attribute. This attribute is an optional attribute that is defined for a group of destinations. In Dell Networking OS, you can assign a COMMUNITY attribute to BGP routers by using an IP community list. After you create an IP community list, you can apply routing decisions to all routers meeting the criteria in the IP community list.

IETF RFC 1997 defines the COMMUNITY attribute and the predefined communities of INTERNET, NO_EXPORT_SUBCONFED, NO_ADVERTISE, and NO_EXPORT. All BGP routes belong to the INTERNET community. In the RFC, the other communities are defined as follows:

- All routes with the NO_EXPORT_SUBCONFED (0xFFFFFF03) community attribute are not sent to CONFED-EBGP or EBGP peers, but are sent to IBGP peers within CONFED-SUB-AS.
- All routes with the NO_ADVERTISE (0xFFFFFF02) community attribute must not be advertised.
- All routes with the NO_EXPORT (0xFFFFFF01) community attribute must not be advertised outside a BGP confederation boundary, but are sent to CONFED-EBGP and IBGP peers.

Dell Networking OS also supports BGP Extended Communities as described in RFC 4360 — BGP Extended Communities Attribute.

To configure an IP community list, use these commands.

1. Create a community list and enter COMMUNITY-LIST mode.

   CONFIGURATION mode

   `ip community-list` *`community-list-name`*
2. Configure a community list by denying or permitting specific community numbers or types of community.

   CONFIG-COMMUNITYLIST mode

   `{deny | permit} {`*`community-number`*` | local-AS | no-advertise | no-export | quote-regexp `*`regular-expression-list`*` | regexp `*`regular-expression`*`}`

   - *`community-number`*: use AA:NN format where AA is the AS number (2 Bytes or 4 Bytes) and NN is a value specific to that autonomous system.
   - `local-AS`: routes with the COMMUNITY attribute of NO_EXPORT_SUBCONFED.
   - `no-advertise`: routes with the COMMUNITY attribute of NO_ADVERTISE.
   - `no-export`: routes with the COMMUNITY attribute of NO_EXPORT.
   - `quote-regexp`: then any number of regular expressions. The software applies all regular expressions in the list.
   - `regexp`: then a regular expression.

Example of the `show ip community-lists` Command

To view the configuration, use the `show config` command in CONFIGURATION COMMUNITY-LIST or CONFIGURATION EXTCOMMUNITY LIST mode or the `show ip {`*`community-lists`*` | `*`extcommunity-list`*`}` command in EXEC Privilege mode.

```
Dell#show ip community-lists
ip community-list standard 1
  deny 701:20
  deny 702:20
  deny 703:20
  deny 704:20
  deny 705:20
  deny 14551:20
  deny 701:112
  deny 702:112
  deny 703:112
  deny 704:112
  deny 705:112
  deny 14551:112
  deny 701:667
  deny 702:667
  deny 703:667
  deny 704:666
  deny 705:666
  deny 14551:666
Dell#
```

Configuring an IP Extended Community List

To configure an IP extended community list, use these commands.

1. Create a extended community list and enter the EXTCOMMUNITY-LIST mode.

   CONFIGURATION mode

   `ip extcommunity-list *extcommunity-list-name*`
2. Two types of extended communities are supported.

   CONFIG-COMMUNITY-LIST mode

   `{permit | deny} {{rt | soo} {*ASN:NN* | *IPADDR:N*} | regex *REGEX-LINE*}`

   Filter routes based on the type of extended communities they carry using one of the following keywords:

   - `rt`: route target.
   - `soo`: route origin or site-of-origin. Support for matching extended communities against regular expression is also supported. Match against a regular expression using the following keyword.
   - `regexp`: regular expression.

Example of the `show ip extcommunity-lists` Command

To set or modify an extended community attribute, use the `set extcommunity {rt | soo} {ASN:NN | IPADDR:NN}` command.

To view the configuration, use the `show config` command in CONFIGURATION COMMUNITY-LIST or CONFIGURATION EXTCOMMUNITY LIST mode or the `show ip {*community-lists* | *extcommunity-list*}` command in EXEC Privilege mode.

```
Dell#show ip community-lists
ip community-list standard 1
  deny 701:20
  deny 702:20
  deny 703:20
  deny 704:20
  deny 705:20
  deny 14551:20
  deny 701:112
  deny 702:112
  deny 703:112
  deny 704:112
  deny 705:112
  deny 14551:112
  deny 701:667
  deny 702:667
  deny 703:667
  deny 704:666
  deny 705:666
  deny 14551:666
Dell#
```

Filtering Routes with Community Lists

To use an IP community list or IP extended community list to filter routes, you must apply a match community filter to a route map and then apply that route map to a BGP neighbor or peer group.

1. Enter the ROUTE-MAP mode and assign a name to a route map.

   CONFIGURATION mode

   ```
   route-map map-name [permit | deny] [sequence-number]
   ```

2. Configure a match filter for all routes meeting the criteria in the IP community or IP extended community list.

   CONFIG-ROUTE-MAP mode

   ```
   match {community community-list-name [exact] | extcommunity extcommunity-
   list-name [exact]}
   ```

3. Return to CONFIGURATION mode.

   CONFIG-ROUTE-MAP mode

   ```
   exit
   ```

4. Enter ROUTER BGP mode.

   CONFIGURATION mode

   ```
   router bgp as-number
   ```

   `AS-number`: 0 to 65535 (2-Byte) or 1 to 4294967295 (4-Byte) or 0.1 to 65535.65535 (Dotted format)

5. Apply the route map to the neighbor or peer group's incoming or outgoing routes.

   CONFIG-ROUTER-BGP mode

   ```
   neighbor {ip-address | peer-group-name} route-map map-name {in | out}
   ```

To view the BGP configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode. To view a route map configuration, use the `show route-map` command in EXEC Privilege mode.

To view which BGP routes meet an IP community or IP extended community list's criteria, use the `show ip bgp {community-list | extcommunity-list}` command in EXEC Privilege mode.

Manipulating the COMMUNITY Attribute

In addition to permitting or denying routes based on the values of the COMMUNITY attributes, you can manipulate the COMMUNITY attribute value and send the COMMUNITY attribute with the route information.

By default, Dell Networking OS does not send the COMMUNITY attribute.

To send the COMMUNITY attribute to BGP neighbors, use the following command.

- Enable the software to send the router's COMMUNITY attribute to the BGP neighbor or peer group specified.

  CONFIG-ROUTER-BGP mode

  ```
  neighbor {ip-address | peer-group-name} send-community
  ```

To view the BGP configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode.

If you want to remove or add a specific COMMUNITY number from a BGP path, you must create a route map with one or both of the following statements in the route map. Then apply that route map to a BGP neighbor or peer group.

1. Enter ROUTE-MAP mode and assign a name to a route map.

   CONFIGURATION mode

   ```
   route-map map-name [permit | deny] [sequence-number]
   ```
2. Configure a set filter to delete all COMMUNITY numbers in the IP community list.

   CONFIG-ROUTE-MAP mode

   ```
   set comm-list community-list-name delete
   ```

   OR

   ```
   set community {community-number | local-as | no-advertise | no-export |
   none}
   ```

   Configure a community list by denying or permitting specific community numbers or types of community.

   - `community-number`: use AA:NN format where AA is the AS number (2 or 4 Bytes) and NN is a value specific to that autonomous system.
   - `local-AS`: routes with the COMMUNITY attribute of NO_EXPORT_SUBCONFED and are not sent to EBGP peers.
   - `no-advertise`: routes with the COMMUNITY attribute of NO_ADVERTISE and are not advertised.
   - `no-export`: routes with the COMMUNITY attribute of NO_EXPORT.
   - `none`: remove the COMMUNITY attribute.
   - `additive`: add the communities to already existing communities.
3. Return to CONFIGURATION mode.

   CONFIG-ROUTE-MAP mode

   ```
   exit
   ```
4. Enter the ROUTER BGP mode.

   CONFIGURATION mode

   ```
   router bgp as-number
   ```
5. Apply the route map to the neighbor or peer group's incoming or outgoing routes.

   CONFIG-ROUTER-BGP mode

   ```
   neighbor {ip-address | peer-group-name} route-map map-name {in | out}
   ```

Example of the `show ip bgp community` Command

To view the BGP configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode. To view a route map configuration, use the `show route-map` command in EXEC Privilege mode.

To view BGP routes matching a certain community number or a pre-defined BGP community, use the `show ip bgp community` command in EXEC Privilege mode.

```
Dell>show ip bgp community
BGP table version is 3762622, local router ID is 10.114.8.48
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

Network            Next Hop  Metric  LocPrf Weight Path
* i 3.0.0.0/8      195.171.0.16      100    0      209 701 80 i
*>i 4.2.49.12/30   195.171.0.16      100    0      209 i
* i 4.21.132.0/23  195.171.0.16      100    0      209 6461 16422 i
*>i 4.24.118.16/30 195.171.0.16      100    0      209 i
*>i 4.24.145.0/30  195.171.0.16      100    0      209 i
*>i 4.24.187.12/30 195.171.0.16      100    0      209 i
*>i 4.24.202.0/30  195.171.0.16      100    0      209 i
*>i 4.25.88.0/30   195.171.0.16      100    0      209 3561 3908 i
*>i 6.1.0.0/16     195.171.0.16      100    0      209 7170 1455 i
*>i 6.2.0.0/22     195.171.0.16      100    0      209 7170 1455 i
*>i 6.3.0.0/18     195.171.0.16      100    0      209 7170 1455 i
*>i 6.4.0.0/16     195.171.0.16      100    0      209 7170 1455 i
*>i 6.5.0.0/19     195.171.0.16      100    0      209 7170 1455 i
*>i 6.8.0.0/20     195.171.0.16      100    0      209 7170 1455 i
*>i 6.9.0.0/20     195.171.0.16      100    0      209 7170 1455 i
*>i 6.10.0.0/15    195.171.0.16      100    0      209 7170 1455 i
*>i 6.14.0.0/15    205.171.0.16      100    0      209 7170 1455 i
*>i 6.133.0.0/21   205.171.0.16      100    0      209 7170 1455 i
*>i 6.151.0.0/16   205.171.0.16      100    0      209 7170 1455 i
--More--
```

Changing MED Attributes

By default, Dell Networking OS uses the MULTI_EXIT_DISC or MED attribute when comparing EBGP paths from the same AS.
To change how the MED attribute is used, enter any or all of the following commands.

- Enable MED comparison in the paths from neighbors with different ASs.

  CONFIG-ROUTER-BGP mode

  ```
  bgp always-compare-med
  ```

  By default, this comparison is not performed.
- Change the bestpath MED selection.

  CONFIG-ROUTER-BGP mode

  ```
  bgp bestpath med {confed | missing-as-best}
  ```

  - `confed`: Chooses the bestpath MED comparison of paths learned from BGP confederations.
  - `missing-as-best`: Treat a path missing an MED as the most preferred one.

To view the nondefault values, use the `show config` command in CONFIGURATION ROUTER BGP mode.

Changing the LOCAL_PREFERENCE Attribute

In Dell Networking OS, you can change the value of the LOCAL_PREFERENCE attribute.
To change the default values of this attribute for all routes received by the router, use the following command.

- Change the LOCAL_PREF value.

CONFIG-ROUTER-BGP mode

```
bgp default local-preference value
```

- `value`: the range is from 0 to 4294967295.

The default is **100**.

To view the BGP configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode or the `show running-config bgp` command in EXEC Privilege mode.

A more flexible method for manipulating the LOCAL_PREF attribute value is to use a route map.

1. Enter the ROUTE-MAP mode and assign a name to a route map.
   CONFIGURATION mode

   ```
   route-map map-name [permit | deny] [sequence-number]
   ```
2. Change LOCAL_PREF value for routes meeting the criteria of this route map.
   CONFIG-ROUTE-MAP mode

   ```
   set local-preference value
   ```
3. Return to CONFIGURATION mode.
   CONFIG-ROUTE-MAP mode

   ```
   exit
   ```
4. Enter ROUTER BGP mode.
   CONFIGURATION mode

   ```
   router bgp as-number
   ```
5. Apply the route map to the neighbor or peer group's incoming or outgoing routes.
   CONFIG-ROUTER-BGP mode

   ```
   neighbor {ip-address | peer-group-name} route-map map-name {in | out}
   ```

To view the BGP configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode. To view a route map configuration, use the `show route-map` command in EXEC Privilege mode.

Changing the NEXT_HOP Attribute

You can change how the NEXT_HOP attribute is used.
To change how the NEXT_HOP attribute is used, enter the first command. To view the BGP configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode or the `show running-config bgp` command in EXEC Privilege mode.

You can also use route maps to change this and other BGP attributes. For example, you can include the second command in a route map to specify the next hop address.

- Disable next hop processing and configure the router as the next hop for a BGP neighbor.
  CONFIG-ROUTER-BGP mode

  ```
  neighbor {ip-address | peer-group-name} next-hop-self
  ```
- Sets the next hop address.
  CONFIG-ROUTE-MAP mode

```
set next-hop ip-address
```

Changing the WEIGHT Attribute

To change how the WEIGHT attribute is used, enter the first command. You can also use route maps to change this and other BGP attributes. For example, you can include the second command in a route map to specify the next hop address.

- Assign a weight to the neighbor connection.

  CONFIG-ROUTER-BGP mode

  ```
  neighbor {ip-address | peer-group-name} weight weight
  ```

  - `weight`: the range is from 0 to 65535.

  The default is **0**.
- Sets weight for the route.

  CONFIG-ROUTE-MAP mode

  ```
  set weight weight
  ```

  - `weight`: the range is from 0 to 65535.

To view BGP configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode or the `show running-config bgp` command in EXEC Privilege mode.

Enabling Multipath

By default, the software allows one path to a destination. You can enable multipath to allow up to 64 parallel paths to a destination.

NOTE: Dell Networking recommends *not* using multipath and add path simultaneously in a route reflector.

To allow more than one path, use the following command.

The `show ip bgp network` command includes multipath information for that network.

- Enable multiple parallel paths.

  CONFIG-ROUTER-BGP mode

  ```
  maximum-paths {ebgp | ibgp} number
  ```

Filtering BGP Routes

Filtering routes allows you to implement BGP policies.
You can use either IP prefix lists, route maps, AS-PATH ACLs or IP community lists (using a route map) to control which routes the BGP neighbor or peer group accepts and advertises. Prefix lists filter routes based on route and prefix length, while AS-Path ACLs filter routes based on the ASN. Route maps can filter and set conditions, change attributes, and assign update policies.

NOTE: Dell Networking OS supports up to 255 characters in a set community statement inside a route map.

NOTE: With Dell Networking OS, you can create inbound and outbound policies. Each of the commands used for filtering has `in` and `out` parameters that you must apply. In Dell Networking OS, the order of preference varies depending on whether the attributes are applied for inbound updates or outbound updates.

For inbound and outbound updates the order of preference is:

- prefix lists (using the `neighbor distribute-list` command)
- AS-PATH ACLs (using the `neighbor filter-list` command)
- route maps (using the `neighbor route-map` command)

Prior to filtering BGP routes, create the prefix list, AS-PATH ACL, or route map.

For configuration information about prefix lists, AS-PATH ACLs, and route maps, refer to Access Control Lists (ACLs).

NOTE: When you configure a new set of BGP policies, to ensure the changes are made, always reset the neighbor or peer group by using the `clear ip bgp` command in EXEC Privilege mode.

To filter routes using prefix lists, use the following commands.

1. Create a prefix list and assign it a name.

   CONFIGURATION mode

   ```
   ip prefix-list prefix-name
   ```
2. Create multiple prefix list filters with a deny or permit action.

   CONFIG-PREFIX LIST mode

   ```
   seq sequence-number {deny | permit} {any | ip-prefix [ge | le] }
   ```

   - `ge`: minimum prefix length to be matched.
   - `le`: maximum prefix length to me matched.

   For information about configuring prefix lists, refer to Access Control Lists (ACLs).
3. Return to CONFIGURATION mode.

   CONFIG-PREFIX LIST mode

   ```
   exit
   ```
4. Enter ROUTER BGP mode.

   CONFIGURATION mode

   ```
   router bgp as-number
   ```
5. Filter routes based on the criteria in the configured prefix list.

   CONFIG-ROUTER-BGP mode

   ```
   neighbor {ip-address | peer-group-name} distribute-list prefix-list-name {in
   | out}
   ```

   Configure the following parameters:

   - *`ip-address`* or *`peer-group-name`*: enter the neighbor's IP address or the peer group's name.
   - *`prefix-list-name`*: enter the name of a configured prefix list.
   - `in`: apply the prefix list to inbound routes.
   - `out`: apply the prefix list to outbound routes.

As a reminder, the following are rules concerning prefix lists:

- If the prefix list contains no filters, all routes are permitted.
- If none of the routes match any of the filters in the prefix list, the route is denied. This action is called an implicit deny. (If you want to forward all routes that do not match the prefix list criteria, you must configure a prefix list filter to permit all routes. For example, you could have the following filter as the last filter in your prefix list permit 0.0.0.0/0 le 32).
- After a route matches a filter, the filter's action is applied. No additional filters are applied to the route.

To view the BGP configuration, use the `show config` command in ROUTER BGP mode. To view a prefix list configuration, use the `show ip prefix-list detail` or `show ip prefix-list summary` commands in EXEC Privilege mode.

Filtering BGP Routes Using Route Maps

To filter routes using a route map, use these commands.

1. Create a route map and assign it a name.

   CONFIGURATION mode

   ```
   route-map map-name [permit | deny] [sequence-number]
   ```
2. Create multiple route map filters with a match or set action.

   CONFIG-ROUTE-MAP mode

   ```
   {match | set}
   ```

   For information about configuring route maps, refer to Access Control Lists (ACLs).
3. Return to CONFIGURATION mode.

   CONFIG-ROUTE-MAP mode

   ```
   exit
   ```
4. Enter ROUTER BGP mode.

   CONFIGURATION mode

   ```
   router bgp as-number
   ```
5. Filter routes based on the criteria in the configured route map.

   CONFIG-ROUTER-BGP mode

   ```
   neighbor {ip-address | peer-group-name} route-map map-name {in | out}
   ```

   Configure the following parameters:

   - *`ip-address`* or *`peer-group-name`*: enter the neighbor's IP address or the peer group's name.
   - *`map-name`*: enter the name of a configured route map.
   - `in`: apply the route map to inbound routes.
   - `out`: apply the route map to outbound routes.

To view the BGP configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode. To view a route map configuration, use the `show route-map` command in EXEC Privilege mode.

Filtering BGP Routes Using AS-PATH Information

To filter routes based on AS-PATH information, use these commands.

1. Create a AS-PATH ACL and assign it a name.
   CONFIGURATION mode
   ```
   ip as-path access-list as-path-name
   ```
2. Create a AS-PATH ACL filter with a deny or permit action.
   AS-PATH ACL mode
   ```
   {deny | permit} as-regular-expression
   ```
3. Return to CONFIGURATION mode.
   AS-PATH ACL
   ```
   exit
   ```
4. Enter ROUTER BGP mode.
   CONFIGURATION mode
   ```
   router bgp as-number
   ```
5. Filter routes based on the criteria in the configured route map.
   CONFIG-ROUTER-BGP mode
   ```
   neighbor {ip-address | peer-group-name} filter-list as-path-name {in | out}
   ```
   Configure the following parameters:
   - *`ip-address`* or *`peer-group-name`*: enter the neighbor's IP address or the peer group's name.
   - *`as-path-name`*: enter the name of a configured AS-PATH ACL.
   - `in`: apply the AS-PATH ACL map to inbound routes.
   - `out`: apply the AS-PATH ACL to outbound routes.

To view which commands are configured, use the `show config` command in CONFIGURATION ROUTER BGP mode and the `show ip as-path-access-list` command in EXEC Privilege mode.

To forward all routes not meeting the AS-PATH ACL criteria, include the **permit .*** filter in your AS-PATH ACL.

Configuring BGP Route Reflectors

BGP route reflectors are intended for ASs with a large mesh; they reduce the amount of BGP control traffic.

NOTE: Dell Networking recommends *not* using multipath and add path simultaneously in a route reflector.

With route reflection configured properly, IBGP routers are not fully meshed within a cluster but all receive routing information.

Configure clusters of routers where one router is a concentration router and the others are clients who receive their updates from the concentration router.

To configure a route reflector, use the following commands.

- Assign an ID to a router reflector cluster.

  CONFIG-ROUTER-BGP mode

  ```
  bgp cluster-id cluster-id
  ```

  You can have multiple clusters in an AS.
- Configure the local router as a route reflector and the neighbor or peer group identified is the route reflector client.

  CONFIG-ROUTER-BGP mode

  ```
  neighbor {ip-address | peer-group-name} route-reflector-client
  ```

When you enable a route reflector, Dell Networking OS automatically enables route reflection to all clients. To disable route reflection between all clients in this reflector, use the `no bgp client-to-client reflection` command in CONFIGURATION ROUTER BGP mode. All clients must be fully meshed before you disable route reflection.
To view a route reflector configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode or the `show running-config bgp` in EXEC Privilege mode.

Aggregating Routes

Dell Networking OS provides multiple ways to aggregate routes in the BGP routing table. At least one specific route of the aggregate must be in the routing table for the configured aggregate to become active.
To aggregate routes, use the following command.

AS_SET includes AS_PATH and community information from the routes included in the aggregated route.

- Assign the IP address and mask of the prefix to be aggregated.

  CONFIG-ROUTER-BGP mode

  ```
  aggregate-address ip-address mask [advertise-map map-name] [as-set]
  [attribute-map map-name] [summary-only] [suppress-map map-name]
  ```

Example of Viewing Aggregated Routes

In the `show ip bgp` command, aggregates contain an 'a' in the first column (shown in bold) and routes suppressed by the aggregate contain an 's' in the first column.

```
Dell#show ip bgp
BGP table version is 0, local router ID is 10.101.15.13
Status codes: s suppressed, d damped, h history, * valid, > best
Path source: I - internal, a - aggregate, c - confed-external, r - redistributed,
n - network
Origin codes: i - IGP, e - EGP, ? - incomplete

Network            Next Hop     Metric LocPrf Weight Path
*> 7.0.0.0/29      10.114.8.33  0             0 18508 ?
*> 7.0.0.0/30      10.114.8.33  0             0 18508 ?
*>a 9.0.0.0/8      192.0.0.0             32768 18508 701 {7018 2686 3786} ?
*> 9.2.0.0/16      10.114.8.33                0 18508 701 i
*> 9.141.128.0/24 10.114.8.33                 0 18508 701 7018 2686 ?
Dell#
```

Configuring BGP Confederations

Another way to organize routers within an AS and reduce the mesh for IBGP peers is to configure BGP confederations.
As with route reflectors, BGP confederations are recommended only for IBGP peering involving many IBGP peering sessions per router. Basically, when you configure BGP confederations, you break the AS into smaller sub-AS, and to those outside your network, the confederations appear as one AS. Within the confederation sub-AS, the IBGP neighbors are fully meshed and the MED, NEXT_HOP, and LOCAL_PREF attributes are maintained between confederations.

To configure BGP confederations, use the following commands.

- Specifies the confederation ID.

  CONFIG-ROUTER-BGP mode

  ```
  bgp confederation identifier as-number
  ```

  – *as-number*: from 0 to 65535 (2 Byte) or from 1 to 4294967295 (4 Byte).
- Specifies which confederation sub-AS are peers.

  CONFIG-ROUTER-BGP mode

  ```
  bgp confederation peers as-number [... as-number]
  ```

  – *as-number*: from 0 to 65535 (2 Byte) or from 1 to 4294967295 (4 Byte).

  All Confederation routers must be either 4 Byte or 2 Byte. You cannot have a mix of router ASN support.

To view the configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode.

Enabling Route Flap Dampening

When EBGP routes become unavailable, they “flap” and the router issues both WITHDRAWN and UPDATE notices.
A flap is when a route:

- is withdrawn
- is readvertised after being withdrawn
- has an attribute change

The constant router reaction to the WITHDRAWN and UPDATE notices causes instability in the BGP process. To minimize this instability, you may configure penalties (a numeric value) for routes that flap. When that penalty value reaches a configured limit, the route is not advertised, even if the route is up. In Dell Networking OS, that penalty value is 1024. As time passes and the route does not flap, the penalty value decrements or is decayed. However, if the route flaps again, it is assigned another penalty.

The penalty value is cumulative and penalty is added under following cases:

- Withdraw
- Readvertise
- Attribute change

When dampening is applied to a route, its path is described by one of the following terms:

- history entry — an entry that stores information on a downed route
- dampened path — a path that is no longer advertised
- penalized path — a path that is assigned a penalty

To configure route flap dampening parameters, set dampening parameters using a route map, clear information on route dampening and return suppressed routes to active state, view statistics on route flapping, or change the path selection from the default mode (deterministic) to non-deterministic, use the following commands.

- Enable route dampening.

  CONFIG-ROUTER-BGP mode

  ```
  bgp dampening [half-life | reuse | suppress max-suppress-time] [route-map
  map-name]
  ```

  Enter the following optional parameters to configure route dampening parameters:

  - *half-life*: the range is from 1 to 45. Number of minutes after which the Penalty is decreased. After the router assigns a Penalty of 1024 to a route, the Penalty is decreased by half after the half-life period expires. The default is **15 minutes**.
  - *reuse*: the range is from 1 to 20000. This number is compared to the flapping route's Penalty value. If the Penalty value is less than the reuse value, the flapping route is once again advertised (or no longer suppressed). Withdrawn routes are removed from history state. The default is **750**.
  - *suppress*: the range is from 1 to 20000. This number is compared to the flapping route's Penalty value. If the Penalty value is greater than the suppress value, the flapping route is no longer advertised (that is, it is suppressed). The default is **2000**.)
  - *max-suppress-time*: the range is from 1 to 255. The maximum number of minutes a route can be suppressed. The default is four times the half-life value. The default is **60 minutes**.
  - `route-map` *map-name*: name of a configured route map. Only match commands in the configured route map are supported. Use this parameter to apply route dampening to selective routes.
- Enter the following optional parameters to configure route dampening.

  CONFIG-ROUTE-MAP mode

  ```
  set dampening half-life reuse suppress max-suppress-time
  ```

  - `half-life`: the range is from 1 to 45. Number of minutes after which the Penalty is decreased. After the router assigns a Penalty of 1024 to a route, the Penalty is decreased by half after the half-life period expires. The default is **15 minutes**.
  - *reuse*: the range is from 1 to 20000. This number is compared to the flapping route's Penalty value. If the Penalty value is less than the reuse value, the flapping route is once again advertised (or no longer suppressed). The default is **750**.
  - *suppress*: the range is from 1 to 20000. This number is compared to the flapping route's Penalty value. If the Penalty value is greater than the suppress value, the flapping route is no longer advertised (that is, it is suppressed). The default is **2000**.
  - *max-suppress-time*: the range is from 1 to 255. The maximum number of minutes a route can be suppressed. The default is four times the half-life value. The default is **60 minutes**.
- Clear all information or only information on a specific route.

  EXEC Privilege

  ```
  clear ip bgp dampening [ip-address mask]
  ```
- View all flap statistics or for specific routes meeting the following criteria.

  EXEC or EXEC Privilege mode

```
show ip bgp flap-statistics [ip-address [mask]] [filter-list as-path-name]
[regexp regular-expression]
```

- `ip-address [mask]`: enter the IP address and mask.
- `filter-list as-path-name`: enter the name of an AS-PATH ACL.
- `regexp regular-expression`: enter a regular express to match on.

By default, the path selection in Dell Networking OS is deterministic, that is, paths are compared irrespective of the order of their arrival. You can change the path selection method to non-deterministic, that is, paths are compared in the order in which they arrived (starting with the most recent). Furthermore, in non-deterministic mode, the software may not compare MED attributes though the paths are from the same AS.

- Change the best path selection method to non-deterministic.

  Change the best path selection method to non-deterministic.

  CONFIG-ROUTER-BGP mode

  ```
  bgp non-deterministic-med
  ```

  NOTE: When you change the best path selection method, path selection for existing paths remains unchanged until you reset it by entering the `clear ip bgp` command in EXEC Privilege mode.

Examples of Configuring a Route and Viewing the Number of Dampened Routes

To view the BGP configuration, use the `show config` command in CONFIGURATION ROUTER BGP mode or the `show running-config bgp` command in EXEC Privilege mode.

The following example shows how to configure values to reuse or restart a route. In the following example, `default = 15` is the set time before the value decrements, `bgp dampening 2 ?` is the set re-advertise value, `bgp dampening 2 2000 ?` is the suppress value, and `bgp dampening 2 2000 3000 ?` is the time to suppress a route. Default values are also shown.

```
Dell(conf-router_bgp)#bgp dampening ?
<1-45> Half-life time for the penalty (default = 15)
route-map Route-map to specify criteria for dampening
<cr>
Dell(conf-router_bgp)#bgp dampening 2 ?
<1-20000>     Value to start reusing a route (default = 750)
Dell(conf-router_bgp)#bgp dampening 2 2000 ?
<1-20000>     Value to start suppressing a route (default = 2000)
Dell(conf-router_bgp)#bgp dampening 2 2000 3000 ?
<1-255>       Maximum duration to suppress a stable route (default = 60)
Dell(conf-router_bgp)#bgp dampening 2 2000 3000 10 ?
route-map     Route-map to specify criteria for dampening
<cr>
```

To view a count of dampened routes, history routes, and penalized routes when you enable route dampening, look at the seventh line of the `show ip bgp summary` command output, as shown in the following example (bold).

```
Dell>show ip bgp summary
BGP router identifier 10.114.8.131, local AS number 65515
BGP table version is 855562, main routing table version 780266
122836 network entrie(s) and 221664 paths using 29697640 bytes of memory
34298 BGP path attribute entrie(s) using 1920688 bytes of memory
29577 BGP AS-PATH entrie(s) using 1384403 bytes of memory
184 BGP community entrie(s) using 7616 bytes of memory
```

Dampening enabled. 0 history paths, 0 dampened paths, 0 penalized paths

```
Neighbor    AS    MsgRcvd MsgSent TblVer  InQ OutQ Up/Down State/PfxRcd
10.114.8.34 18508 82883   79977   780266  0   2 00:38:51   118904
10.114.8.33 18508 117265  25069   780266  0   20 00:38:50  102759
Dell>
```

To view which routes are dampened (non-active), use the `show ip bgp dampened-routes` command in EXEC Privilege mode.

Changing BGP Timers

To configure BGP timers, use either or both of the following commands.

Timer values configured with the `neighbor timers` command override the timer values configured with the `timers bgp` command.

When two neighbors, configured with different `keepalive` and `holdtime` values, negotiate for new values, the resulting values are as follows:

- the lower of the `holdtime` values is the new `holdtime` value, and
- whichever is the lower value; one-third of the new `holdtime` value, or the configured `keepalive` value is the new `keepalive` value.

- Configure timer values for a BGP neighbor or peer group.

  CONFIG-ROUTER-BGP mode

  ```
  neighbors {ip-address | peer-group-name} timers keepalive holdtime
  ```

  - *`keepalive`*: the range is from 1 to 65535. Time interval, in seconds, between keepalive messages sent to the neighbor routers. The default is **60 seconds**.
  - *`holdtime`*: the range is from 3 to 65536. Time interval, in seconds, between the last keepalive message and declaring the router dead. The default is **180 seconds**.
- Configure timer values for all neighbors.

  CONFIG-ROUTER-BGP mode

  ```
  timers bgp keepalive holdtime
  ```

  - *`keepalive`*: the range is from 1 to 65535. Time interval, in seconds, between keepalive messages sent to the neighbor routers. The default is **60 seconds**.
  - *`holdtime`*: the range is from 3 to 65536. Time interval, in seconds, between the last keepalive message and declaring the router dead. The default is **180 seconds**.

To view non-default values, use the `show config` command in CONFIGURATION ROUTER BGP mode or the `show running-config bgp` command in EXEC Privilege mode.

Enabling BGP Neighbor Soft-Reconfiguration

BGP soft-reconfiguration allows for faster and easier route changing.

Changing routing policies typically requires a reset of BGP sessions (the TCP connection) for the policies to take effect. Such resets cause undue interruption to traffic due to hard reset of the BGP cache and the time it takes to re-establish the session. BGP soft reconfig allows for policies to be applied to a session without clearing the BGP Session. Soft-reconfig can be done on a per-neighbor basis and can either be inbound or outbound.

BGP soft-reconfiguration clears the policies without resetting the TCP connection.

To reset a BGP connection using BGP soft reconfiguration, use the `clear ip bgp` command in EXEC Privilege mode at the system prompt.

When you enable soft-reconfiguration for a neighbor and you execute the `clear ip bgp soft in` command, the update database stored in the router is replayed and updates are reevaluated. With this command, the replay and update process is triggered only if a route-refresh request is not negotiated with the peer. If the request is indeed negotiated (after execution of `clear ip bgp soft in`), BGP sends a route-refresh request to the neighbor and receives all of the peer's updates.

To use soft reconfiguration (or soft reset) without preconfiguration, both BGP peers must support the soft route refresh capability, which is advertised in the open message sent when the peers establish a TCP session.

To determine whether a BGP router supports this capability, use the `show ip bgp neighbors` command. If a router supports the route refresh capability, the following message displays: `Received route refresh capability from peer.`

If you specify a BGP peer group by using the *`peer-group-name`* argument, all members of the peer group inherit the characteristic configured with this command.

- Clear all information or only specific details.
  EXEC Privilege mode

  ```
  clear ip bgp {* | neighbor-address | AS Numbers | ipv4 | peer-group-name}
  [soft [in | out]]
  ```

  - `*`: Clears all peers.
  - `neighbor-address`: Clears the neighbor with this IP address.
  - `AS Numbers`: Peers' AS numbers to be cleared.
  - `ipv4`: Clears information for the IPv4 address family.
  - `peer-group-name`: Clears all members of the specified peer group.
- Enable soft-reconfiguration for the BGP neighbor specified.
  CONFIG-ROUTER-BGP mode

  ```
  neighbor {ip-address | peer-group-name} soft-reconfiguration inbound
  ```

  BGP stores all the updates received by the neighbor but does not reset the peer-session.

  Entering this command starts the storage of updates, which is required to do inbound soft reconfiguration. Outbound BGP soft reconfiguration does not require inbound soft reconfiguration to be enabled.

Example of Soft-Reconfigration of a BGP Neighbor

The example enables inbound soft reconfiguration for the neighbor 10.108.1.1. All updates received from this neighbor are stored unmodified, regardless of the inbound policy. When inbound soft reconfiguration is done later, the stored information is used to generate a new set of inbound updates.

```
Dell>router bgp 100
   neighbor 10.108.1.1 remote-as 200
   neighbor 10.108.1.1 soft-reconfiguration inbound
```

Route Map Continue

The BGP route map continue feature, `continue [sequence-number]`, (in ROUTE-MAP mode) allows movement from one route-map entry to a specific route-map entry (the sequence number).

If you do not specify a sequence number, the continue feature moves to the next sequence number (also known as an "implied continue"). If a match clause exists, the continue feature executes only after a successful match occurs. If there are no successful matches, continue is ignored.

Match a Clause with a Continue Clause

The continue feature can exist without a match clause.

Without a match clause, the continue clause executes and jumps to the specified route-map entry. With a match clause and a continue clause, the match clause executes first and the continue clause next in a specified route map entry. The continue clause launches only after a successful match. The behavior is:

- A successful match with a continue clause—the route map executes the set clauses and then goes to the specified route map entry after execution of the continue clause.
- If the next route map entry contains a continue clause, the route map executes the continue clause if a successful match occurs.
- If the next route map entry does not contain a continue clause, the route map evaluates normally. If a match does not occur, the route map does not continue and falls-through to the next sequence number, if one exists

Set a Clause with a Continue Clause

If the route-map entry contains sets with the continue clause, the set actions operation is performed first followed by the continue clause jump to the specified route map entry.

- If a set actions operation occurs in the first route map entry and then the same set action occurs with a different value in a subsequent route map entry, the last set of actions overrides the previous set of actions with the same `set` command.
- If the `set community additive` and `set as-path prepend` commands are configured, the communities and AS numbers are prepended.

Enabling MBGP Configurations

Multiprotocol BGP (MBGP) is an enhanced BGP that carries IP multicast routes. BGP carries two sets of routes: one set for unicast routing and one set for multicast routing. The routes associated with multicast routing are used by the protocol independent multicast (PIM) to build data distribution trees.

MBGP for IPv4 multicast is supported on the S6000 platform.

Dell Networking OS MBGP is implemented per RFC 1858. You can enable the MBGP feature per router and/or per peer/peer-group.

The default is **IPv4 Unicast** routes.

When you configure a peer to support IPv4 multicast, Dell Networking OS takes the following actions:

- Send a capacity advertisement to the peer in the BGP Open message specifying IPv4 multicast as a supported AFI/SAFI (Subsequent Address Family Identifier).
- If the corresponding capability is received in the peer's Open message, BGP marks the peer as supporting the AFI/SAFI.

- When exchanging updates with the peer, BGP sends and receives IPv4 multicast routes if the peer is marked as supporting that AFI/SAFI.
- Exchange of IPv4 multicast route information occurs through the use of two new attributes called MP_REACH_NLRI and MP_UNREACH_NLRI, for feasible and withdrawn routes, respectively.
- If the peer has not been activated in any AFI/SAFI, the peer remains in Idle state.

Most Dell Networking OS BGP IPv4 unicast commands are extended to support the IPv4 multicast RIB using extra options to the command. For a detailed description of the MBGP commands, refer to the *Dell Networking OS Command Line Interface Reference Guide*.

- Enables support for the IPv4 multicast family on the BGP node.

  CONFIG-ROUTER-BGP mode

  ```
  address family ipv4 multicast
  ```
- Enable IPv4 multicast support on a BGP neighbor/peer group.

  CONFIG-ROUTER-BGP-AF (Address Family) mode

  ```
  neighbor [ip-address | peer-group-name] activate
  ```

BGP Regular Expression Optimization

Dell Networking OS optimizes processing time when using regular expressions by caching and re-using regular expression evaluated results, at the expense of some memory in RP1 processor.

BGP policies that contain regular expressions to match against as-paths and communities might take a lot of CPU processing time, thus affect BGP routing convergence. Also, `show bgp` commands that get filtered through regular expressions can to take a lot of CPU cycles, especially when the database is large.

This feature is turned on by default. If necessary, use the `bgp regex-eval-optz-disable` command in CONFIGURATION ROUTER BGP mode to disable it.

Debugging BGP

To enable BGP debugging, use any of the following commands.

- View all information about BGP, including BGP events, keepalives, notifications, and updates.

  EXEC Privilege mode

  ```
  debug ip bgp [ip-address | peer-group peer-group-name] [in | out]
  ```
- View information about BGP route being dampened.

  EXEC Privilege mode

  ```
  debug ip bgp dampening [in | out]
  ```
- View information about local BGP state changes and other BGP events.

  EXEC Privilege mode

  ```
  debug ip bgp [ip-address | peer-group peer-group-name] events [in | out]
  ```
- View information about BGP KEEPALIVE messages.

  EXEC Privilege mode

  ```
  debug ip bgp [ip-address | peer-group peer-group-name] keepalive [in | out]
  ```
- View information about BGP notifications received from or sent to neighbors.

EXEC Privilege mode

`debug ip bgp [`*`ip-address`*` | peer-group `*`peer-group-name`*`] notifications [in | out]`

- View information about BGP updates and filter by prefix name.

  EXEC Privilege mode

  `debug ip bgp [`*`ip-address`*` | peer-group `*`peer-group-name`*`] updates [in | out] [prefix-list name]`
- Enable soft-reconfiguration debug.

  EXEC Privilege mode

  `debug ip bgp {`*`ip-address`*` | `*`peer-group-name`*`} soft-reconfiguration`

  To enhance debugging of soft reconfig, use the `bgp soft-reconfig-backup` command only when route-refresh is not negotiated to avoid the peer from resending messages.

  In-BGP is shown using the `show ip protocols` command.

Dell Networking OS displays debug messages on the console. To view which debugging commands are enabled, use the `show debugging` command in EXEC Privilege mode.

To disable a specific `debug` command, use the keyword `no` then the `debug` command. For example, to disable debugging of BGP updates, use `no debug ip bgp updates` command.

To disable all BGP debugging, use the `no debug ip bgp` command.

To disable all debugging, use the `undebug all` command.

Storing Last and Bad PDUs

Dell Networking OS stores the last notification sent/received and the last bad protocol data unit (PDU) received on a per peer basis. The last bad PDU is the one that causes a notification to be issued.

In the following example, the last seven lines shown in bold are the last PDUs.

Example of the `show ip bgp neighbor` Command to View Last and Bad PDUs

```
Dell(conf-router_bgp)#do show ip bgp neighbors 1.1.1.2

BGP neighbor is 1.1.1.2, remote AS 2, external link
  BGP version 4, remote router ID 2.4.0.1
  BGP state ESTABLISHED, in this state for 00:00:01
  Last read 00:00:00, last write 00:00:01
  Hold time is 90, keepalive interval is 30 seconds
  Received 1404 messages, 0 in queue
    3 opens, 1 notifications, 1394 updates
    6 keepalives, 0 route refresh requests
  Sent 48 messages, 0 in queue
    3 opens, 2 notifications, 0 updates
    43 keepalives, 0 route refresh requests
  Minimum time between advertisement runs is 30 seconds
  Minimum time before advertisements start is 0 seconds

Capabilities received from neighbor for IPv4 Unicast :
  MULTIPROTO_EXT(1)
  ROUTE_REFRESH(2)
  CISCO_ROUTE_REFRESH(128)
```

```
Capabilities advertised to neighbor for IPv4 Unicast :
  MULTIPROTO_EXT(1)
  ROUTE_REFRESH(2)
  CISCO_ROUTE_REFRESH(128)
For address family: IPv4 Unicast
BGP table version 1395, neighbor version 1394
Prefixes accepted 1 (consume 4 bytes), 0 withdrawn by peer
Prefixes advertised 0, rejected 0, 0 withdrawn from peer

Connections established 3; dropped 2
Last reset 00:00:12, due to Missing well known attribute

Notification History
  'UPDATE error/Missing well-known attr' Sent : 1 Recv: 0
  'Connection Reset' Sent : 1 Recv: 0

   Last notification (len 21) sent 00:26:02 ago
  ffffffff ffffffff ffffffff ffffffff 00160303 03010000
 Last notification (len 21) received 00:26:20 ago
  ffffffff ffffffff ffffffff ffffffff 00150306 00000000
 Last PDU (len 41) received 00:26:02 ago that caused notification to be issued
  ffffffff ffffffff ffffffff ffffffff 00290200 00000e01 02040201 00024003 04141414 0218c0a8
  01000000
Local host: 1.1.1.1, Local port: 179
Foreign host: 1.1.1.2, Foreign port: 41758
```

Capturing PDUs

To capture incoming and outgoing PDUs on a per-peer basis, use the `capture bgp-pdu neighbor direction` command. To disable capturing, use the `no capture bgp-pdu neighbor direction` command.

The buffer size supports a maximum value between 40 MB (the default) and 100 MB. The capture buffers are cyclic and reaching the limit prompts the system to overwrite the oldest PDUs when new ones are received for a given neighbor or direction. Setting the buffer size to a value lower than the current maximum, might cause captured PDUs to be freed to set the new limit.

NOTE: Memory on RP1 is not pre-allocated and is allocated only when a PDU needs to be captured.

The buffers storing the PDU free memory when:

- BGP is disabled.
- A neighbor is unconfigured.
- The `clear ip bgp` command is issued.
- New PDU are captured and there is no more space to store them.
- The max buffer size is reduced. (This may cause PDUs to be cleared depending on the buffer space consumed and the new limit.)

Examples of the `show capture bgp-pdu neighbor` Command

To change the maximum buffer size, use the `capture bgp-pdu max-buffer-size` command. To view the captured PDUs, use the `show capture bgp-pdu neighbor` command.

```
Dell#show capture bgp-pdu neighbor 20.20.20.2

Incoming packet capture enabled for BGP neighbor 20.20.20.2
Available buffer size 40958758, 26 packet(s) captured using 680 bytes
  PDU[1] : len 101, captured 00:34:51 ago
    ffffffff ffffffff ffffffff ffffffff 00650100 00000013 00000000 00000000
419ef06c 00000000
```

```
    00000000 00000000 00000000 00000000 0181a1e4 0181a25c 41af92c0 00000000
00000000 00000000
    00000000 00000001 0181a1e4 0181a25c 41af9400 00000000
  PDU[2] : len 19, captured 00:34:51 ago
    ffffffff ffffffff ffffffff ffffffff 00130400
  PDU[3] : len 19, captured 00:34:51 ago
    ffffffff ffffffff ffffffff ffffffff 00130400
  PDU[4] : len 19, captured 00:34:22 ago
    ffffffff ffffffff ffffffff ffffffff 00130400
[. . .]

Outgoing packet capture enabled for BGP neighbor 20.20.20.2
Available buffer size 40958758, 27 packet(s) captured using 562 bytes
  PDU[1] : len 41, captured 00:34:52 ago
    ffffffff ffffffff ffffffff ffffffff 00290104 000100b4 14141401 0c020a01
04000100 01020080
    00000000
  PDU[2] : len 19, captured 00:34:51 ago
    ffffffff ffffffff ffffffff ffffffff 00130400
  PDU[3] : len 19, captured 00:34:50 ago
    ffffffff ffffffff ffffffff ffffffff 00130400
  PDU[4] : len 19, captured 00:34:20 ago
    ffffffff ffffffff ffffffff ffffffff 00130400
[. . .]
```

The following example shows how to view space requirements for storing all the PDUs. With full internet feed (205K) captured, approximately 11.8MB is required to store all of the PDUs.

```
Dell(conf-router_bgp)#do show capture bgp-pdu neighbor 172.30.1.250

Incoming packet capture enabled for BGP neighbor 172.30.1.250
Available buffer size 29165743, 192991 packet(s) captured using 11794257 bytes
  [. . .]

Dell(conf-router_bgp)#do sho ip bg s
BGP router identifier 172.30.1.56, local AS number 65056
BGP table version is 313511, main routing table version 313511
207896 network entrie(s) and 207896 paths using 42364576 bytes of memory
59913 BGP path attribute entrie(s) using 2875872 bytes of memory
59910 BGP AS-PATH entrie(s) using 2679698 bytes of memory
3 BGP community entrie(s) using 81 bytes of memory

Neighbor     AS     MsgRcvd  MsgSent  TblVer InQ OutQ Up/Down State/Pfx
1.1.1.2      2      17       18966    0      0   0    00:08:19 Active
172.30.1.250 18508  243295   25       313511 0   0    00:12:46 207896
```

PDU Counters

Dell Networking OS supports additional counters for various types of PDUs sent and received from neighbors.

These are seen in the output of the `show ip bgp neighbor` command.

Sample Configurations

The following example configurations show how to enable BGP and set up some peer groups. These examples are not comprehensive directions. They are intended to give you some guidance with typical configurations.

To support your own IP addresses, interfaces, names, and so on, you can copy and paste from these examples to your CLI. Be sure that you make the necessary changes.

The following illustration shows the configurations described on the following examples. These configurations show how to create BGP areas using physical and virtual links. They include setting up the interfaces and peers groups with each other.

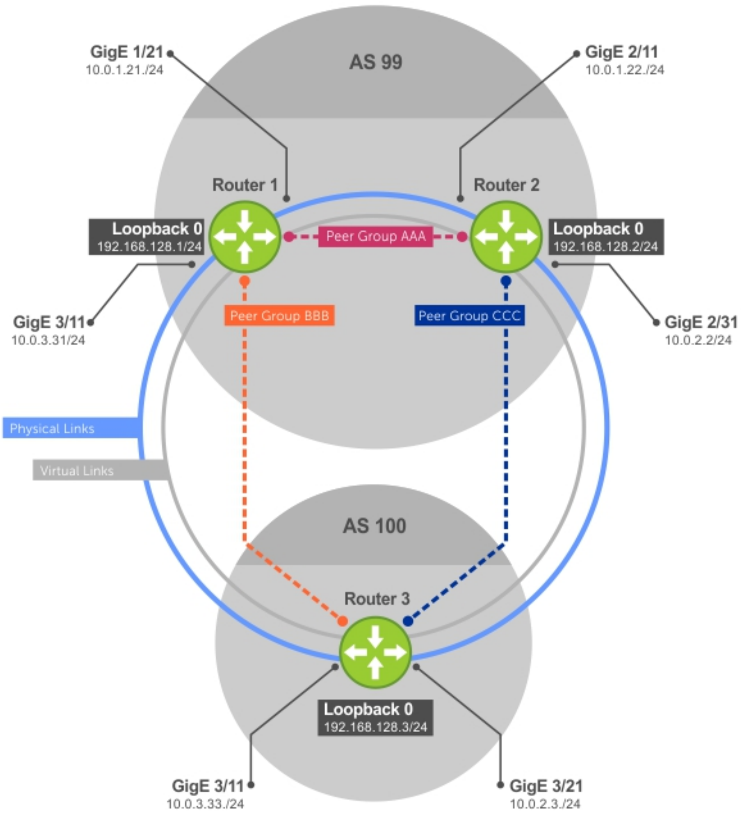

GigE 1/21
10.0.1.21./24
AS 99
GigE 2/11
10.0.1.22./24
Router 1
Router 2
Loopback 0
192.168.128.1/24
Peer Group AAA
Loopback 0
192.168.128.2/24
GigE 3/11
10.0.3.31/24
Peer Group BBB
Peer Group CCC
GigE 2/31
10.0.2.2/24
Physical Links
Virtual Links
AS 100
Router 3
Loopback 0
192.168.128.3/24
GigE 3/11
10.0.3.33./24
GigE 3/21
10.0.2.3./24

Figure 24. Sample Configurations

Example of Enabling BGP (Router 1)

```
R1# conf
R1(conf)#int loop 0
R1(conf-if-lo-0)#ip address 192.168.128.1/24
R1(conf-if-lo-0)#no shutdown
R1(conf-if-lo-0)#show config
!
interface Loopback 0
ip address 192.168.128.1/24
```

```
no shutdown
R1(conf-if-lo-0)#int te 1/21
R1(conf-if-te-1/21)#ip address 10.0.1.21/24
R1(conf-if-te-1/21)#no shutdown
R1(conf-if-te-1/21)#show config
!
interface TengigabitEthernet 1/21
ip address 10.0.1.21/24
no shutdown
R1(conf-if-te-1/21)#int te 1/31
R1(conf-if-te-1/31)#ip address 10.0.3.31/24
R1(conf-if-te-1/31)#no shutdown
R1(conf-if-te-1/31)#show config
!
interface TengigabitEthernet 1/31
ip address 10.0.3.31/24
no shutdown
R1(conf-if-te-1/31)#router bgp 99
R1(conf-router_bgp)#network 192.168.128.0/24
R1(conf-router_bgp)#neighbor 192.168.128.2 remote 99
R1(conf-router_bgp)#neighbor 192.168.128.2 no shut
R1(conf-router_bgp)#neighbor 192.168.128.2 update-source loop 0
R1(conf-router_bgp)#neighbor 192.168.128.3 remote 100
R1(conf-router_bgp)#neighbor 192.168.128.3 no shut
R1(conf-router_bgp)#neighbor 192.168.128.3 update-source loop 0
R1(conf-router_bgp)#show config
!
router bgp 99
network 192.168.128.0/24
neighbor 192.168.128.2 remote-as 99
neighbor 192.168.128.2 update-source Loopback 0
neighbor 192.168.128.2 no shutdown
neighbor 192.168.128.3 remote-as 100
neighbor 192.168.128.3 update-source Loopback 0
neighbor 192 168 128 3 no shutdown
```

Example of Enabling BGP (Router 2)

```
R2# conf
R2(conf)#int loop 0
R2(conf-if-lo-0)#ip address 192.168.128.2/24
R2(conf-if-lo-0)#no shutdown
R2(conf-if-lo-0)#show config
!
interface Loopback 0
ip address 192.168.128.2/24
no shutdown
R2(conf-if-lo-0)#int te 2/11
R2(conf-if-te-2/11)#ip address 10.0.1.22/24
R2(conf-if-te-2/11)#no shutdown
R2(conf-if-te-2/11)#show config
!
interface TengigabitEthernet 2/11
ip address 10.0.1.22/24
no shutdown
R2(conf-if-te-2/11)#int te 2/31
R2(conf-if-te-2/31)#ip address 10.0.2.2/24
R2(conf-if-te-2/31)#no shutdown
R2(conf-if-te-2/31)#show config
!
interface TengigabitEthernet 2/31
ip address 10.0.2.2/24
no shutdown
R2(conf-if-te-2/31)#
R2(conf-if-te-2/31)#router bgp 99
```

```
R2(conf-router_bgp)#network 192.168.128.0/24
R2(conf-router_bgp)#neighbor 192.168.128.1 remote 99
R2(conf-router_bgp)#neighbor 192.168.128.1 no shut
R2(conf-router_bgp)#neighbor 192.168.128.1 update-source loop 0
R2(conf-router_bgp)#neighbor 192.168.128.3 remote 100
R2(conf-router_bgp)#neighbor 192.168.128.3 no shut
R2(conf-router_bgp)#neighbor 192.168.128.3 update loop 0
R2(conf-router_bgp)#show config
!
router bgp 99
bgp router-id 192.168.128.2
network 192.168.128.0/24
```

Example of Enabling BGP (Router 3)

```
R3# conf
R3(conf)#
R3(conf)#int loop 0
R3(conf-if-lo-0)#ip address 192.168.128.3/24
R3(conf-if-lo-0)#no shutdown
R3(conf-if-lo-0)#show config
!
interface Loopback 0
ip address 192.168.128.3/24
no shutdown
R3(conf-if-lo-0)#int te 3/11
R3(conf-if-te-3/11)#ip address 10.0.3.33/24
R3(conf-if-te-3/11)#no shutdown
R3(conf-if-te-3/11)#show config
!
interface TengigabitEthernet 3/11
ip address 10.0.3.33/24
no shutdown
R3(conf-if-lo-0)#int te 3/21
R3(conf-if-te-3/21)#ip address 10.0.2.3/24
R3(conf-if-te-3/21)#no shutdown
R3(conf-if-te-3/21)#show config
!
interface TengigabitEthernet 3/21
ip address 10.0.2.3/24
no shutdown
R3(conf-if-te-3/21)#
R3(conf-if-te-3/21)#router bgp 100
R3(conf-router_bgp)#show config
!
router bgp 100
R3(conf-router_bgp)#network 192.168.128.0/24
R3(conf-router_bgp)#neighbor 192.168.128.1 remote 99
R3(conf-router_bgp)#neighbor 192.168.128.1 no shut
R3(conf-router_bgp)#neighbor 192.168.128.1 update-source loop 0
R3(conf-router_bgp)#neighbor 192.168.128.2 remote 99
R3(conf-router_bgp)#neighbor 192.168.128.2 no shut
R3(conf-router_bgp)#neighbor 192.168.128.2 update loop 0
R3(conf-router_bgp)#show config
```

Example of Enabling Peer Groups (Router 1)

```
conf
R1(conf)#router bgp 99
R1(conf-router_bgp)# network 192.168.128.0/24
R1(conf-router_bgp)# neighbor AAA peer-group
R1(conf-router_bgp)# neighbor AAA no shutdown
R1(conf-router_bgp)# neighbor BBB peer-group
R1(conf-router_bgp)# neighbor BBB no shutdown
R1(conf-router_bgp)# neighbor 192.168.128.2 peer-group AAA
```

```
R1(conf-router_bgp)# neighbor 192.168.128.3 peer-group BBB
R1(conf-router_bgp)#
R1(conf-router_bgp)#show config
!
router bgp 99
network 192.168.128.0/24
neighbor AAA peer-group
neighbor AAA no shutdown
neighbor BBB peer-group
neighbor BBB no shutdown
neighbor 192.168.128.2 remote-as 99
neighbor 192.168.128.2 peer-group AAA
neighbor 192.168.128.2 update-source Loopback 0
neighbor 192.168.128.2 no shutdown
neighbor 192.168.128.3 remote-as 100
neighbor 192.168.128.3 peer-group BBB
neighbor 192.168.128.3 update-source Loopback 0
neighbor 192.168.128.3 no shutdown
R1#
R1#show ip bgp summary
BGP router identifier 192.168.128.1, local AS number 99
BGP table version is 1, main routing table version 1
1 network entrie(s) using 132 bytes of memory
3 paths using 204 bytes of memory
BGP-RIB over all using 207 bytes of memory
2 BGP path attribute entrie(s) using 96 bytes of memory
2 BGP AS-PATH entrie(s) using 74 bytes of memory
2 neighbor(s) using 8672 bytes of memory
Neighbor AS MsgRcvd MsgSent TblVer InQ OutQ Up/Down State/Pfx
192.168.128.2 99 23 24 1 0 (0) 00:00:17 1Capabilities received from neighbor
for IPv4 Unicast :
MULTIPROTO_EXT(1)
ROUTE_REFRESH(2)
CISCO_ROUTE_REFRESH(128)
Capabilities advertised to neighbor for IPv4 Unicast :
MULTIPROTO_EXT(1)
ROUTE_REFRESH(2)
CISCO_ROUTE_REFRESH(128)
Update source set to Loopback 0
Peer active in peer-group outbound optimization
For address family: IPv4 Unicast
BGP table version 1, neighbor version 1
Prefixes accepted 1 (consume 4 bytes), withdrawn 0 by peer
Prefixes advertised 1, denied 0, withdrawn 0 from peer
Connections established 2; dropped 1
Last reset 00:00:57, due to user reset
Notification History
'Connection Reset' Sent : 1 Recv: 0
Last notification (len 21) sent 00:00:57 ago
ffffffff ffffffff ffffffff ffffffff 00150306 00000000
Local host: 192.168.128.1, Local port: 179
Foreign host: 192.168.128.2, Foreign port: 65464
BGP neighbor is 192.168.128.3, remote AS 100, external link
Member of peer-group BBB for session parameters
BGP version 4, remote router ID 192.168.128.3
BGP state ESTABLISHED, in this state for 00:00:37
Last read 00:00:36, last write 00:00:36
Hold time is 180, keepalive interval is 60 seconds
Received 30 messages, 0 in queue
4 opens, 2 notifications, 4 updates
20 keepalives, 0 route refresh requests
Sent 29 messages, 0 in queue
4 opens, 1 notifications, 4 updates
20 keepalives, 0 route refresh requests
```

```
Minimum time between advertisement runs is 30 seconds
Minimum time before advertisements start is 0 seconds
```

Example of Enabling Peer Groups (Router 2)

```
R2#conf
R2(conf)#router bgp 99
R2(conf-router_bgp)# neighbor CCC peer-group
R2(conf-router_bgp)# neighbor CC no shutdown
R2(conf-router_bgp)# neighbor BBB peer-group
R2(conf-router_bgp)# neighbor BBB no shutdown
R2(conf-router_bgp)# neighbor 192.168.128.1 peer AAA
R2(conf-router_bgp)# neighbor 192.168.128.1 no shut
R2(conf-router_bgp)# neighbor 192.168.128.3 peer BBB
R2(conf-router_bgp)# neighbor 192.168.128.3 no shut
R2(conf-router_bgp)#show conf
!
router bgp 99
network 192.168.128.0/24
neighbor AAA peer-group
neighbor AAA no shutdown
neighbor BBB peer-group
neighbor BBB no shutdown
neighbor 192.168.128.1 remote-as 99
neighbor 192.168.128.1 peer-group CCC
neighbor 192.168.128.1 update-source Loopback 0
neighbor 192.168.128.1 no shutdown
neighbor 192.168.128.3 remote-as 100
neighbor 192.168.128.3 peer-group BBB
neighbor 192.168.128.3 update-source Loopback 0
neighbor 192.168.128.3 no shutdown
R2(conf-router_bgp)#end
R2#
R2#show ip bgp summary
BGP router identifier 192.168.128.2, local AS number 99
BGP table version is 2, main routing table version 2
1 network entrie(s) using 132 bytes of memory
3 paths using 204 bytes of memory
BGP-RIB over all using 207 bytes of memory
2 BGP path attribute entrie(s) using 128 bytes of memory
2 BGP AS-PATH entrie(s) using 90 bytes of memory
2 neighbor(s) using 9216 bytes of memory
Neighbor AS MsgRcvd MsgSent TblVer InQ OutQ Up/Down State/Pfx
192.168.128.1 99 140 136 2 0 (0) 00:11:24 1
192.168.128.3 100 138 140 2 0 (0) 00:18:31 1
```

Example of Enabling Peer Groups (Router 3)

```
R3#conf
R3(conf)#router bgp 100
R3(conf-router_bgp)# neighbor AAA peer-group
R3(conf-router_bgp)# neighbor AAA no shutdown
R3(conf-router_bgp)# neighbor CCC peer-group
R3(conf-router_bgp)# neighbor CCC no shutdown
R3(conf-router_bgp)# neighbor 192.168.128.2 peer-group BBB
R3(conf-router_bgp)# neighbor 192.168.128.2 no shutdown
R3(conf-router_bgp)# neighbor 192.168.128.1 peer-group BBB
R3(conf-router_bgp)# neighbor 192.168.128.1 no shutdown
R3(conf-router_bgp)#
R3(conf-router_bgp)#end
R3#show ip bgp summary
BGP router identifier 192.168.128.3, local AS number 100
BGP table version is 1, main routing table version 1
1 network entrie(s) using 132 bytes of memory
3 paths using 204 bytes of memory
```

```
BGP-RIB over all using 207 bytes of memory
2 BGP path attribute entrie(s) using 128 bytes of memory
2 BGP AS-PATH entrie(s) using 90 bytes of memory
2 neighbor(s) using 9216 bytes of memory
Neighbor AS MsgRcvd MsgSent TblVer InQ OutQ Up/Down State/Pfx
192.168.128.1 99 93 99 1 0 (0) 00:00:15 1
192.168.128.2 99 122 120 1 0 (0) 00:00:11 1
R3#show ip bgp neighbor
BGP neighbor is 192.168.128.1, remote AS 99, external link
Member of peer-group BBB for session parameters
BGP version 4, remote router ID 192.168.128.1
BGP state ESTABLISHED, in this state for 00:00:21
Last read 00:00:09, last write 00:00:08
Hold time is 180, keepalive interval is 60 seconds
Received 93 messages, 0 in queue
5 opens, 0 notifications, 5 updates
83 keepalives, 0 route refresh requests
Sent 99 messages, 0 in queue
5 opens, 4 notifications, 5 updates
85 keepalives, 0 route refresh requestsCapabilities received from neighbor for
IPv4 Unicast :
MULTIPROTO_EXT(1)
ROUTE_REFRESH(2)
CISCO_ROUTE_REFRESH(128)
Capabilities advertised to neighbor for IPv4 Unicast :
MULTIPROTO_EXT(1)
ROUTE_REFRESH(2)
CISCO_ROUTE_REFRESH(128)
Update source set to Loopback 0
Peer active in peer-group outbound optimization
For address family: IPv4 Unicast
BGP table version 2, neighbor version 2
Prefixes accepted 1 (consume 4 bytes), withdrawn 0 by peer
Prefixes advertised 1, denied 0, withdrawn 0 from peer
Connections established 6; dropped 5
Last reset 00:12:01, due to Closed by neighbor
Notification History
'HOLD error/Timer expired' Sent : 1 Recv: 0
'Connection Reset' Sent : 2 Recv: 2
Last notification (len 21) received 00:12:01 ago
ffffffff ffffffff ffffffff ffffffff 00150306 00000000
Local host: 192.168.128.2, Local port: 65464
Foreign host: 192.168.128.1, Foreign port: 179
BGP neighbor is 192.168.128.3, remote AS 100, external link
Member of peer-group BBB for session parameters
BGP version 4, remote router ID 192.168.128.3
BGP state ESTABLISHED, in this state for 00:18:51
Last read 00:00:45, last write 00:00:44
Hold time is 180, keepalive interval is 60 seconds
Received 138 messages, 0 in queue
7 opens, 2 notifications, 7 updates
122 keepalives, 0 route refresh requests
Sent 140 messages, 0 in queue
```

9

Content Addressable Memory (CAM)

Content addressable memory (CAM) is supported on the S6000 platform.

CAM is a type of memory that stores information in the form of a lookup table. On Dell Networking systems, CAM stores Layer 2 and Layer 3 forwarding information, access-lists (ACLs), flows, and routing policies.

CAM Allocation

The user configurable CAM allocations feature is available on the S6000 platform.

CAM Allocation for Ingress

To allocate the space for regions such has L2 ingress ACL, IPV4 ingress ACL, IPV6 ingress ACL, IPV4 QOS, L2 QOS, PBR, VRF ACL etc on the S-Series by using the `cam-acl` command in **CONFIGURATION** mode.

The CAM space is allotted in Field Processor (FP) blocks. The total space allocated must equal 13 FP blocks.

NOTE: There are 16 FP blocks, but the system flow requires three blocks that cannot be reallocated.

The following table lists the default CAM allocation settings.

Table 8. Default Cam Allocation Settings

CAM Allocation	Setting
L2Acl	6
IPV4Acl	4
Ipv6Acl	0
Ipv4Qos	2
L2Qos	1
L2PT	0
IpMacAcl	0
VmanQos	0
VmanDualQos	0
EcfmAcl	0
FcoeAcl	0
iscsiOptAcl	0
ipv4pbr	0
vrfv4Acl	0

CAM Allocation	Setting
Openflow	0
fedgovacl	0

The following additional CAM allocation settings are supported on the S6000, S4810 or S4820T platforms only.

Table 9. Additional Default CAM Allocation Settings

Additional CAM Allocation	Setting
FCoE ACL (fcoeacl)	0
ISCSI Opt ACL (iscsioptacl)	0

The `ipv6acl` and `vman-dual-qos` allocations must be entered as a factor of 2 (2, 4, 6, 8, 10). All other profile allocations can use either even or odd numbered ranges.

You must save the new CAM settings to the startup-config (`write-mem` or `copy run start`) then reload the system for the new settings to take effect.

CAM Allocation for Ingress

Use the cam-acl-egress command to allocate the space for egress L2, IPV4 and IPV6 ACL. The total number of available FP blocks is 4. Allocate atleast one group of L2ACL and IPV4 ACL.

```
Dell(conf)#do show cam-acl-egress

-- Chassis Egress Cam ACL --
             Current Settings(in block sizes)
                   1 block = 256 entries
L2Acl        :         1
Ipv4Acl      :         1
Ipv6Acl      :         2

-- Stack unit 0 --
           Current Settings(in block sizes)
L2Acl        :         1
Ipv4Acl      :         1
Ipv6Acl      :         2

-- Stack unit 7 --
           Current Settings(in block sizes)
L2Acl        :         1
Ipv4Acl      :         1
Ipv6Acl      :         2

Dell(conf)#
```

1. Select a cam-acl action.

   CONFIGURATION mode

   ```
   cam-acl [default | l2acl]
   ```

   NOTE: Selecting default resets the CAM entries to the default settings. Select `l2acl` to allocate the desired space for all other regions.

2. Enter the number of FP blocks for each region.

EXEC Privilege mode

```
cam-acl {default | l2acl number ipv4acl number ipv6acl number ipv4qos number
l2qos number l2pt number ipmacacl number vman-qos | vman-dual-qos number
ecfmacl number ipv4pbr number openflow number | fcoe number iscsioptacl
number [vrfv4acl number]
```

NOTE: If the allocation values are not entered for the CAM regions, the value is 0.

3. Execute the write memory, verify that the new settings will be written to the CAM on the next boot.

   EXEC Privilege mode

   ```
   show cam-acl
   ```

4. Reload the system.

   EXEC Privilege mode

   ```
   reload
   ```

Test CAM Usage

The `test cam-usage` command is supported on the S6000 platform.

Use this command to determine whether sufficient CAM space is available to enable a service-policy. Create a Class Map with all required ACL rules, then execute the `test cam-usage` command in Privilege mode to verify the actual CAM space required. The Status column in the command output indicates whether or not the policy can be enabled.

Example of the `test cam-usage` Command

```
Dell#test cam-usage service-policy input test-cam-usage stack-unit 7 po 0

Stack-Unit | Portpipe | CAM Partition | Available CAM | Estimated CAM per Port
| Status
--------------------------------------------------------------------------------
----------
   7       |    0      | IPv4Flow      | 192           | 3
| Allowed (64)
Dell#
```

View CAM Profiles

To view the current CAM profile for the chassis and each component, use the `show cam-profile` command.

This command also shows the profile that is loaded after the next chassis or component reload.

Example of the `show cam-profile` Command

```
Dell#show cam-profile

-- Chassis CAM Profile --

CamSize         : 18-Meg
                : Current Settings : Next Boot
Profile Name    : Default          : Default
L2FIB           : 32K entries      : 32K entries
L2ACL           : 1K entries       : 1K entries
IPv4FIB         : 256K entries     : 256K entries
IPv4ACL         : 12K entries      : 12K entries
```

```
IPv4Flow       : 24K entries      : 24K entries
EgL2ACL        : 1K entries       : 1K entries
EgIPv4ACL      : 1K entries       : 1K entries
Reserved       : 8K entries       : 8K entries
FIB          : 0 entries      : 0 entries
ACL          : 0 entries      : 0 entries
Flow         : 0 entries      : 0 entries
EgACL        : 0 entries      : 0 entries
MicroCode Name : Default          : Default
--More--
```

To view brief output of the `show cam-profile` command, use the `summary` option.

The `show running-config cam-profile` command shows the current profile and microcode.

NOTE: If you select the CAM profile from CONFIGURATION mode, the output of this command does not reflect any changes until you save the running-configuration and reload the chassis.

Example of `show running-config cam-profile` Command

```
Dell#show running-config cam-profile
!
cam-profile default microcode default

Dell#
```

View CAM-ACL Settings

The `show cam-acl` command is supported on the S6000 platform.

This`show cam-acl` command shows the cam-acl setting that will be loaded after the next reload.

Example of Viewing CAM-ACL Settings

```
Dell(conf)#do show cam-acl

-- Chassis Cam ACL --
Current Settings(in block sizes)   Next Boot(in block sizes)
                   1 block = 128 entries
L2Acl        :         6                            4
Ipv4Acl      :         4                            2
Ipv6Acl      :         0                            0
Ipv4Qos      :         2                            2
L2Qos        :         1                            1
L2PT         :         0                            0
IpMacAcl     :         0                            0
VmanQos      :         0                            0
VmanDualQos  :         0                            0
EcfmAcl      :         0                            0
FcoeAcl      :         0                            0
iscsiOptAcl  :         0                            0
ipv4pbr      :         0                            2
vrfv4Acl     :         0                            2
Openflow     :         0                            0
fedgovacl    :         0                            0

-- Stack unit 0 --
 Current Settings(in block sizes)  Next Boot(in block sizes)
                   1 block = 128 entries
L2Acl        :         6                            4
Ipv4Acl      :         4                            2
Ipv6Acl      :         0                            0
Ipv4Qos      :         2                            2
L2Qos        :         1                            1
```

```
L2PT         :          0                             0
IpMacAcl     :          0                             0
VmanQos      :          0                             0
VmanDualQos  :          0                             0
EcfmAcl      :          0                             0
FcoeAcl      :          0                             0
iscsiOptAcl  :          0                             0
ipv4pbr      :          0                             2
vrfv4Acl     :          0                             2
Openflow     :          0                             0
fedgovacl    :          0                             0

Dell(conf)#
```

Example of Viewing CAM-ACL Settings (S6000)

NOTE: If you change the cam-acl setting from the **CONFIGURATION** mode, the output of this command does not reflect any changes until you save the running-configuration and reload the chassis.

The default values for the `show cam-acl` command for the S6000 are:

```
Dell#show cam-acl

-- Chassis Cam ACL --
             Current Settings(in block sizes)
                   1 block = 128 entries
L2Acl        :          6
Ipv4Acl      :          4
Ipv6Acl      :          0
Ipv4Qos      :          2
L2Qos        :          1
L2PT         :          0
IpMacAcl     :          0
VmanQos      :          0
VmanDualQos  :          0
EcfmAcl      :          0
FcoeAcl      :          0
iscsiOptAcl  :          0
ipv4pbr      :          0
vrfv4Acl     :          0
Openflow     :          0
fedgovacl    :          0

-- Stack unit 0 --
           Current Settings(in block sizes)
                   1 block = 128 entries
L2Acl        :          6
Ipv4Acl      :          4
Ipv6Acl      :          0
Ipv4Qos      :          2
L2Qos        :          1
L2PT         :          0
IpMacAcl     :          0
VmanQos      :          0
VmanDualQos  :          0
EcfmAcl      :          0
FcoeAcl      :          0
iscsiOptAcl  :          0
ipv4pbr      :          0
vrfv4Acl     :          0
Openflow     :          0
fedgovacl    :          0
```

```
-- Stack unit 7 --
           Current Settings(in block sizes)
                   1 block = 128 entries
L2Acl        :         6
Ipv4Acl      :         4
Ipv6Acl      :         0
Ipv4Qos      :         2
L2Qos        :         1
L2PT         :         0
IpMacAcl     :         0
VmanQos      :         0
VmanDualQos  :         0
EcfmAcl      :         0
FcoeAcl      :         0
iscsiOptAcl  :         0
ipv4pbr      :         0
vrfv4Acl     :         0
Openflow     :         0
fedgovacl    :         0

Dell#
```

View CAM Usage

View the amount of CAM space available, used, and remaining in each ACL partition using the `show cam-usage` command from EXEC Privilege mode.

Example of the `show cam-usage` Command

```
Dell#show cam-usage
Stackunit|Portpipe| CAM Partition   | Total CAM   |  Used CAM   |Available CAM
========|========|=================|=============|=============|==============
    0      0       | IN-L3 ACL     |       512    |         1    |       511
        |          |  IN-V6 ACL    |         0    |         0    |         0
        |          | IN-L2 ACL     |       768    |         0    |       768
        |          | OUT-L3 ACL    |       158    |         5    |       153
        |          | OUT-V6 ACL    |       158    |          0   |       158
        |          | OUT-L2 ACL    |       206    |          7   |       199
    7   |   0      | IN-L3 ACL     |       512    |          1   |       511
        |          | IN-V6 ACL     |         0    |           0 |         0
        |          | IN-L2 ACL     |        768   |           0 |       768
        |          | OUT-L3 ACL    |       158    |           5 |       153
        |          | OUT-V6 ACL    |       158    |           0 |       158
        |          | OUT-L2 ACL    |       206    |           7 |       199
Codes: * - cam usage is above 90%.
Dell#
```

CAM Optimization

CAM optimization is supported on the S6000 platform.

When you enable this command, if a Policy Map containing classification rules (ACL and/or dscp/ ip-precedence rules) is applied to more than one physical interface on the same port-pipe, only a single copy of the policy is written (only 1 FP entry is used).

Troubleshoot CAM Profiling

The following section describes CAM profiling troubleshooting.

CAM Profile Mismatches

The CAM profile on all cards must match the system profile. In most cases, the system corrects mismatches by copying the correct profile to the card, and rebooting the card. If three resets do not bring up the card, or if the system is running an Dell Networking OS version prior to version 6.3.1.1, the system presents an error message. In this case, manually adjust the CAM configuration on the card to match the system configuration.

Dell Networking recommends the following to prevent mismatches:

- Use the eg-default CAM profile in a chassis that has only EG Series line cards. If this profile is used in a chassis with non-EG line cards, the non-EG line cards enter a problem state.
- Before moving a card to a new chassis, change the CAM profile on a card to match the new system profile.
- After installing a secondary RPM into a chassis, copy the running-configuration to the startup-configuration.
- Change to the default profile if downgrading to and Dell Networking OS version earlier than 6.3.1.1.
- Use the CONFIGURATION mode commands so that the profile is change throughout the system.
- Use the EXEC Privilege mode commands to match the profile of a component to the profile of the target system.

QoS CAM Region Limitation

To store QoS service policies, the default CAM profile allocates a partition within the IPv4Flow region. If the QoS CAM space is exceeded, a message similar to the following displays.

```
     %EX2YD:12 %DIFFSERV-2-DSA_QOS_CAM_INSTALL_FAILED: Not enough space in L3
Cam(PolicyQos) for class 2 (Gi 12/20) entries on portpipe 1 for linecard 12
     %EX2YD:12 %DIFFSERV-2-
     DSA_QOS_CAM_INSTALL_FAILED: Not enough space in L3 Cam(PolicyQos) for
class 5 (Gi 12/
22) entries on portpipe 1 for linecard 12
```

If you exceed the QoS CAM space, follow these steps.

1. Verify that you have configured a CAM profile that allocates 24 K entries to the IPv4 system flow region.
2. Allocate more entries in the IPv4Flow region to QoS.

Dell Networking OS supports the ability to view the actual CAM usage before applying a service-policy. The `test cam-usage service-policy` command provides this test framework. For more information, refer to Pre-Calculating Available QoS CAM Space.

Syslog Error When the Table is Full

In the Dell Networking OS, the table full condition is displayed as CAM full only for LPM. But now the LPM is split into two tables. There are two syslog errors that are displayed:

1. /65 to /128 Table full.
2. 0/0 – 0/64 Table full.

A table-full error message is displayed once the number of entries is crossed the table size. Table-full message is generated only once when it crosses the threshold. For subsequent addition of entries, the table-full message is not recorded you clear the table-full message. The table-full message is cleared internally when the number of entries is less than the table size.

Syslog Warning Upon 90 Percent Utilization of CAM

CAM utilization includes both the L3_DEFIP and L3_DEFIP_PAIR_128 table entries to calculate the utilization.

Syslog Warning for Discrepancies Between Configured Extended Prefixes

An error message is displayed if the number of extended prefix entries is different from the configured value during bootup.

10

Control Plane Policing (CoPP)

Control plane policing (CoPP) is supported on the S6000 platform.

Control plane policing (CoPP) uses access control list (ACL) rules and quality of service (QoS) policies to create filters for a system’s control plane. That filter prevents traffic not specifically identified as legitimate from reaching the system control plane, rate-limits, traffic to an acceptable level.

CoPP increases security on the system by protecting the routing processor from unnecessary or DoS traffic, giving priority to important control plane and management traffic. CoPP uses a dedicated control plane configuration through the ACL and QoS command line interfaces (CLIs) to provide filtering and rate-limiting capabilities for the control plane packets.

The following illustration shows an example of the difference between having CoPP implemented and not having CoPP implemented.

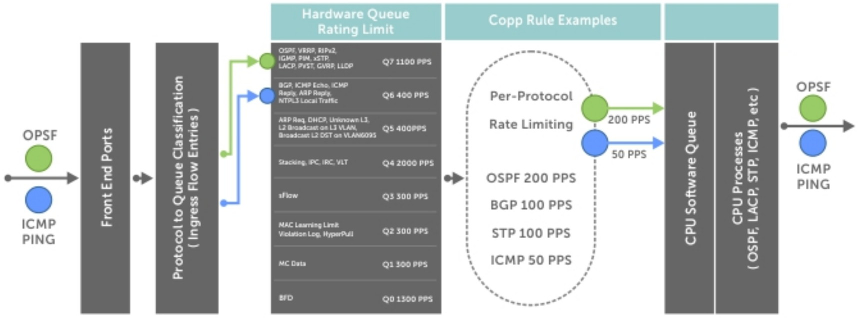

Hardware Queue Rating Limit
Copp Rule Examples
OPSF
ICMP PING
Front End Ports
Protocol to Queue Classification (Ingress Flow Entries)
OSPF, VRRP, RIPv2, IGMP, PIM, xSTP, LACP, PVST, GVRP, LLDP
Q7 1100 PPS
BGP, ICMP Echo, ICMP Reply, ARP Reply, NTPL3 Local Traffic
Q6 400 PPS
ARP Req, DHCP, Unknown L3, L2 Broadcast on L3 VLAN, Broadcast L2 DST on VLAN6095
Q5 400PPS
Stacking, IPC, IRC, VLT
Q4 2000 PPS
sFlow
Q3 300 PPS
MAC Learning Limit Violation Log, HyperPull
Q2 300 PPS
MC Data
Q1 300 PPS
BFD
Q0 1300 PPS
Per-Protocol
Rate Limiting
OSPF 200 PPS
BGP 100 PPS
STP 100 PPS
ICMP 50 PPS
200 PPS
50 PPS
CPU Software Queue
CPU Processes (OSPF, LACP, STP, ICMP, etc)
OPSF
ICMP PING

Figure 25. Control Plane Policing

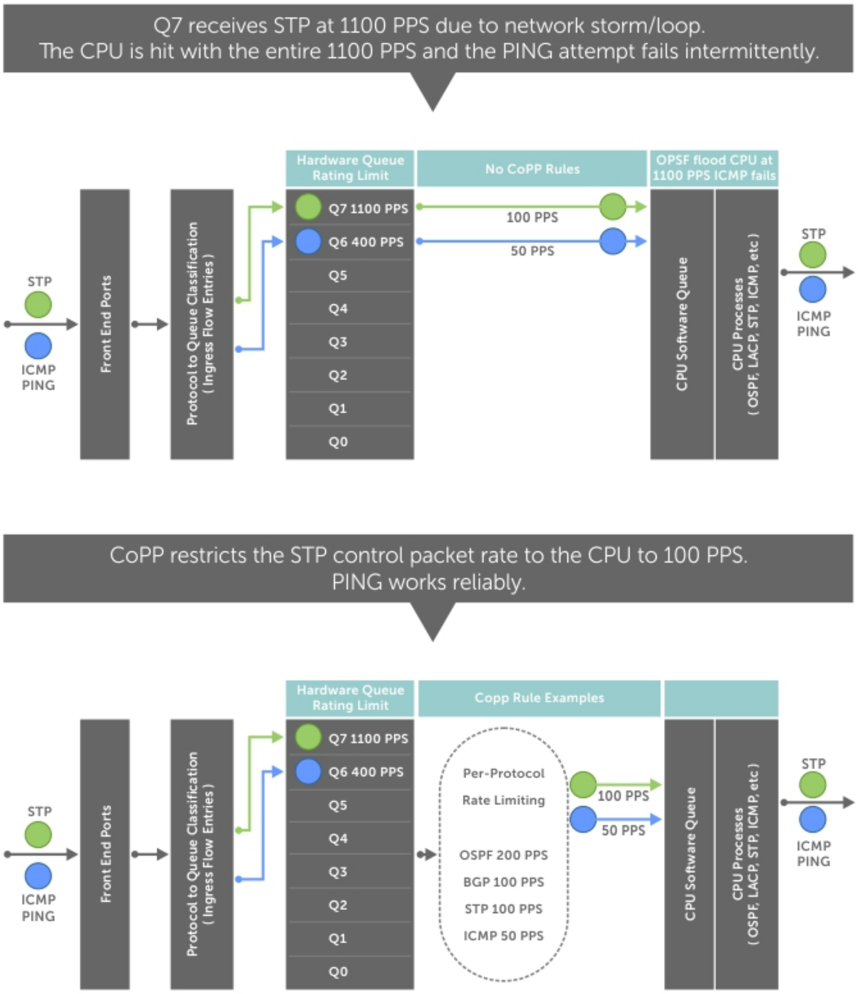

Q7 receives STP at 1100 PPS due to network storm/loop.
The CPU is hit with the entire 1100 PPS and the PING attempt fails intermittently.
Hardware Queue Rating Limit
No CoPP Rules
OPSF flood CPU at 1100 PPS ICMP fails
STP
ICMP PING
Front End Ports
Protocol to Queue Classification (Ingress Flow Entries)
Q7 1100 PPS
Q6 400 PPS
Q5
Q4
Q3
Q2
Q1
Q0
100 PPS
50 PPS
CPU Software Queue
CPU Processes (OSPF, LACP, STP, ICMP, etc)
STP
ICMP PING
CoPP restricts the STP control packet rate to the CPU to 100 PPS.
PING works reliably.
Hardware Queue Rating Limit
Copp Rule Examples
STP
ICMP PING
Front End Ports
Protocol to Queue Classification (Ingress Flow Entries)
Q7 1100 PPS
Q6 400 PPS
Q5
Q4
Q3
Q2
Q1
Q0
Per-Protocol
Rate Limiting
OSPF 200 PPS
BGP 100 PPS
STP 100 PPS
ICMP 50 PPS
100 PPS
50 PPS
CPU Software Queue
CPU Processes (OSPF, LACP, STP, ICMP, etc)
STP
ICMP PING

Figure 26. CoPP Implemented Versus CoPP Not Implemented

Configure Control Plane Policing

For example, border gateway protocol (BGP) and internet control message protocol (ICMP) share same queue (Q6); Q6 has 400 PPS of bandwidth by default. The desired rate of ICMP is 100 PPS and the remaining 300 PPS is assigned to BGP. If ICMP packets come at 400 PPS, BGP packets may be dropped though ICMP packets are rate-limited to 100 PPS. You can solve this by increasing Q6 bandwidth to 700 PPS to allow both ICMP and BGP packets and then applying per-flow CoPP for ICMP and BGP packets. The setting of this Q6 bandwidth is dependent on the incoming traffic for the set of protocols sharing the same queue. If you are not aware of the incoming protocol traffic rate, you cannot set the required queue rate limit value. You must complete queue bandwidth tuning carefully because the system cannot open up to handle any rate, including traffic coming at the line rate.

CoPP policies are assigned on a per-protocol or a per-queue basis, and are assigned in CONTROL-PLANE mode to each port-pipe.

CoPP policies are configured by creating extended ACL rules and specifying rate-limits through QoS policies. The ACLs and QoS policies are assigned as service-policies.

Configuring CoPP for Protocols

This section lists the commands necessary to create and enable the service-policies for CoPP.
For complete information about creating ACLs and QoS rules, refer to Access Control Lists (ACLs) and Quality of Service (QoS).

The basics for creating a CoPP service policy are to create a Layer 2, Layer 3, and/or an IPv6 ACL rule for the desired protocol type. Then, create a QoS input policy to rate-limit the protocol traffics according to the ACL. The ACL and QoS policies are finally assigned to a control-plane service policy for each port-pipe.

1. Create a Layer 2 extended ACL for control-plane traffic policing for a particular protocol.

   CONFIGURATION mode

   ```
   mac access-list extended name cpu-qos permit {arp | frrp | gvrp | isis |
   lacp | lldp | stp}
   ```
2. Create a Layer 3 extended ACL for control-plane traffic policing for a particular protocol.

   CONFIGURATION mode

   ```
   ip access-list extended name cpu-qos permit {bgp | dhcp | dhcp-relay | ftp |
   icmp | igmp | msdp | ntp | ospf | pim | ip | ssh | telnet | vrrp}
   ```
3. Create an IPv6 ACL for control-plane traffic policing for a particular protocol.

   CONFIGURATION mode

   ```
   ipv6 access-list name cpu-qos permit {bgp | icmp | vrrp}
   ```
4. Create a QoS input policy for the router and assign the policing.

   CONFIGURATION mode

   ```
   qos-policy-input name cpu-qos rate-police
   ```
5. Create a QoS class map to differentiate the control-plane traffic and assign to an ACL.

   CONFIGURATION mode

   ```
   class-map match-any name cpu-qos match {ip | mac | ipv6} access-group name
   ```
6. Create a QoS input policy map to match to the class-map and qos-policy for each desired protocol.

   CONFIGURATION mode

   ```
   policy-map-input name cpu-qos class-map name qos-policy name
   ```
7. Enter Control Plane mode.

   CONFIGURATION mode

   ```
   control-plane-cpuqos
   ```
8. Assign the protocol based the service policy on the control plane. Enabling this command on a port-pipe automatically enables the ACL and QoS rules creates with the `cpu-qos` keyword.

   CONTROL-PLANE mode

   ```
   service-policy rate-limit-protocols
   ```

Examples of Configuring CoPP for Different Protocols

The following example shows creating the IP/IPv6/MAC extended ACL.

```
Dell(conf)#ip access-list extended ospf cpu-qos
Dell(conf-ip-acl-cpuqos)#permit ospf
Dell(conf-ip-acl-cpuqos)#exit

Dell(conf)#ip access-list extended bgp cpu-qos
Dell(conf-ip-acl-cpuqos)#permit bgp
Dell(conf-ip-acl-cpuqos)#exit

Dell(conf)#mac access-list extended lacp cpu-qos
Dell(conf-mac-acl-cpuqos)#permit lacp
Dell(conf-mac-acl-cpuqos)#exit

Dell(conf)#ipv6 access-list ipv6-icmp cpu-qos
Dell(conf-ipv6-acl-cpuqos)#permit icmp
Dell(conf-ipv6-acl-cpuqos)#exit

Dell(conf)#ipv6 access-list ipv6-vrrp cpu-qos
Dell(conf-ipv6-acl-cpuqos)#permit vrrp
Dell(conf-ipv6-acl-cpuqos)#exit
```

The following example shows creating the QoS input policy.

```
Dell(conf)#qos-policy-in rate_limit_200k cpu-qos
Dell(conf-in-qos-policy-cpuqos)#rate-police 200 40 peak 500 40
Dell(conf-in-qos-policy-cpuqos)#exit

Dell(conf)#qos-policy-in rate_limit_400k cpu-qos
Dell(conf-in-qos-policy-cpuqos)#rate-police 400 50 peak 600 50
Dell(conf-in-qos-policy-cpuqos)#exit

Dell(conf)#qos-policy-in rate_limit_500k cpu-qos
Dell(conf-in-qos-policy-cpuqos)#rate-police 500 50 peak 1000 50
Dell(conf-in-qos-policy-cpuqos)#exit
```

The following example shows creating the QoS class map.

```
Dell(conf)#class-map match-any class_ospf cpu-qos
Dell(conf-class-map-cpuqos)#match ip access-group ospf
Dell(conf-class-map-cpuqos)#exit

Dell(conf)#class-map match-any class_bgp cpu-qos
Dell(conf-class-map-cpuqos)#match ip access-group bgp
Dell(conf-class-map-cpuqos)#exit

Dell(conf)#class-map match-any class_lacp cpu-qos
Dell(conf-class-map-cpuqos)#match mac access-group lacp
Dell(conf-class-map-cpuqos)#exit

Dell(conf)#class-map match-any class-ipv6-icmp cpu-qos
Dell(conf-class-map-cpuqos)#match ipv6 access-group ipv6-icmp
Dell(conf-class-map-cpuqos)#exit
```

The following example shows matching the QoS class map to the QoS policy.

```
Dell(conf)#policy-map-input egressFP_rate_policy cpu-qos
Dell(conf-policy-map-in-cpuqos)#class-map class_ospf qos-policy rate_limit_500k
Dell(conf-policy-map-in-cpuqos)#class-map class_bgp qos-policy rate_limit_400k
Dell(conf-policy-map-in-cpuqos)#class-map class_lacp qos-policy rate_limit_200k
Dell(conf-policy-map-in-cpuqos)#class-map class-ipv6 qos-policy rate_limit_200k
Dell(conf-policy-map-in-cpuqos)#exit
```

The following example shows creating the control plane service policy.

```
Dell(conf)#control-plane-cpuqos
Dell(conf-control-cpuqos)#service-policy rate-limit-protocols
egressFP_rate_policy
Dell(conf-control-cpuqos)#exit
```

Configuring CoPP for CPU Queues

Controlling traffic on the CPU queues does not require ACL rules, but does require QoS policies. CoPP for CPU queues converts the input rate from kbps to pps, assuming 64 bytes is the average packet size, and applies that rate to the corresponding queue. Consequently, 1 kbps is roughly equivalent to 2 pps.

The basics for creating a CoPP service policy is to create QoS policies for the desired CPU bound queue and associate it with a particular rate-limit. The QoS policies are assigned to a control-plane service policy for each port-pipe.

1. Create a QoS input policy for the router and assign the policing.

   CONFIGURATION mode

   ```
   qos-policy-input name cpu-qos
   ```
2. Create an input policy-map to assign the QoS policy to the desired service queues.l.

   CONFIGURATION mode

   ```
   policy-map--input name cpu-qos service-queue 0 qos-policy name
   ```
3. Enter Control Plane mode.

   CONFIGURATION mode

   ```
   control-plane-cpuqos
   ```
4. Assign a CPU queue-based service policy on the control plane in cpu-qos mode. Enabling this command sets the queue rates according to those configured.

   CONTROL-PLANE mode

   ```
   service-policy rate-limit-cpu-queues name
   ```

Examples of Configuring CoPP for CPU Queues

The following example shows creating the QoS policy.

```
Dell#conf
Dell(conf)#qos-policy-input cpuq_1
Dell(conf-qos-policy-in)#rate-police 3000 40 peak 500 40
Dell(conf-qos-policy-in)#exit

Dell(conf)#qos-policy-input cpuq_2
Dell(conf-qos-policy-in)#rate-police 5000 80 peak 600 50
Dell(conf-qos-policy-in)#exit
```

The following example shows assigning the QoS policy to the queues.

```
Dell(conf)#policy-map-input cpuq_rate_policy cpu-qos
Dell(conf-qos-policy-in)#service-queue 5 qos-policy cpuq_1
Dell(conf-qos-policy-in)#service-queue 6 qos-policy cpuq_2
Dell(conf-qos-policy-in)#service-queue 7 qos-policy cpuq_1
```

The following example shows creating the control plane service policy.

```
Dell#conf
Dell(conf)#control-plane
Dell(conf-control-plane)#service-policy rate-limit-cpu-queues cpuq_rate_policy
```

Show Commands

The following section describes the CoPP show commands.

To view the rates for each queue, use the `show cpu-queue rate cp` command.

Example of Viewing Queue Rates

```
Dell#show cpu-queue rate cp
Service-Queue  Rate (PPS)
-------------- -----------
Q0             1300
Q1             300
Q2             300
Q3             300
Q4             2000
Q5             400
Q6             400
Q7             1100
Dell#
```

Example of Viewing Queue Mapping

To view the queue mapping for each configured protocol, use the `show ip protocol-queue-mapping` command.

```
Dell#show ip protocol-queue-mapping
Protocol    Src-Port Dst-Port TcpFlag Queue EgPort Rate (kbps)
--------    -------- -------- ------- ----- ------ -----------
TCP (BGP)    any/179 179/any   _      Q6    CP      100
UDP (DHCP)   67/68   68/67     _      Q6/Q5 CP      _
UDP (DHCP-R) 67      67        _      Q6    CP      _
TCP (FTP)    any     21        _      Q6    CP      _
ICMP         any     any       _      Q6    CP      _
IGMP         any     any       _      Q7    CP      _
TCP (MSDP)   any/639 639/any   _      Q6    CP      _
UDP (NTP)    any     123       _      Q6    CP      _
OSPF         any     any       _      Q7    CP      _
PIM          any     any       _      Q7    CP      _
UDP (RIP)    any     520       _      Q7    CP      _
TCP (SSH)    any     22        _      Q6    CP      _
TCP (TELNET) any     23        _      Q6    CP      _
VRRP         any     any       _      Q7    CP      _
Dell#
```

To view the queue mapping for the MAC protocols, use the `show mac protocol-queue-mapping` command.

Example of Viewing Queue Mapping for MAC Protocols

```
Dell#show mac protocol-queue-mapping
Protocol Destination Mac      EtherType Queue EgPort Rate (kbps)
-------- ----------------   ----------- ----- ------ -----------
ARP      any                  0x0806   Q5/Q6   CP     _
FRRP     01:01:e8:00:00:10/11 any      Q7      CP     _
LACP     01:80:c2:00:00:02    0x8809   Q7      CP     _
LLDP     any                  0x88cc   Q7      CP     _
```

```
GVRP     01:80:c2:00:00:21    any      Q7      CP     _
STP      01:80:c2:00:00:00    any      Q7      CP     _
ISIS     01:80:c2:00:00:14/15 any      Q7      CP     _
         09:00:2b:00:00:04/05 any      Q7      CP

Dell#
```

To view the queue mapping for IPv6 protocols, use the `show ipv6 protocol-queue-mapping` command.

Example of Viewing Queue Mapping for IPv6 Protocols

```
Dell#show ipv6 protocol-queue-mapping
Protocol   Src-Port Dst-Port TcpFlag Queue EgPort Rate (kbps)
--------   -------- -------- ------- ----- ------ -----------
TCP (BGP)  any/179  179/any   _      Q6     CP     _
ICMP       any      any       _      Q6     CP     _
VRRP       any      any       _      Q7     CP     _

Dell#
```

11

Data Center Bridging (DCB)

Data center bridging (DCB) is supported on the S6000 platform.

NOTE:

Ethernet Enhancements in Data Center Bridging

The following section describes DCB.

The system supports loading two *DCB_Config* files: *FCoE_DCB_Config* and *iSCSI_DCB_Config*.

These files are located in the root directory flash:/CONFIG_TEMPLATE. After copying the configuration files to the startup config and reloading the system.

The S6000 supports the following DCB features:

- Data center bridging exchange protocol (DCBx)
- Priority-based flow control (PFC)
- Enhanced transmission selection (ETS)

DCB refers to a set of IEEE Ethernet enhancements that provide data centers with a single, robust, converged network to support multiple traffic types, including local area network (LAN), server, and storage traffic. Through network consolidation, DCB results in reduced operational cost, simplified management, and easy scalability by avoiding the need to deploy separate application-specific networks.

For example, instead of deploying an Ethernet network for LAN traffic, include additional storage area networks (SANs) to ensure lossless Fibre Channel traffic, and a separate InfiniBand network for high-performance inter-processor computing within server clusters, only one DCB-enabled network is required in a data center. The Dell Networking switches that support a unified fabric and consolidate multiple network infrastructures use a single input/output (I/O) device called a converged network adapter (CNA).

A CNA is a computer input/output device that combines the functionality of a host bus adapter (HBA) with a network interface controller (NIC). Multiple adapters on different devices for several traffic types are no longer required.

Data center bridging satisfies the needs of the following types of data center traffic in a unified fabric:

LAN traffic	LAN traffic consists of many flows that are insensitive to latency requirements, while certain applications, such as streaming video, are more sensitive to latency. Ethernet functions as a best-effort network that may drop packets in the case of network congestion. IP networks rely on transport protocols (for example, TCP) for reliable data transmission with the associated cost of greater processing overhead and performance impact LAN traffic consists of a large number of flows that are generally insensitive to latency requirements, while certain applications, such as streaming video, are more sensitive to latency. Ethernet functions as a best-effort

	network that may drop packets in case of network congestion. IP networks rely on transport protocols (for example, TCP) for reliable data transmission with the associated cost of greater processing overhead and performance impact.
Storage traffic	Storage traffic based on Fibre Channel media uses the Small Computer System Interface (SCSI) protocol for data transfer. This traffic typically consists of large data packets with a payload of 2K bytes that cannot recover from frame loss. To successfully transport storage traffic, data center Ethernet must provide no-drop service with lossless links.
InterProcess Communication (IPC) traffic	InterProcess Communication (IPC) traffic within high-performance computing clusters to share information. Server traffic is extremely sensitive to latency requirements.

To ensure lossless delivery and latency-sensitive scheduling of storage and service traffic and I/O convergence of LAN, storage, and server traffic over a unified fabric, IEEE data center bridging adds the following extensions to a classical Ethernet network:

- 802.1Qbb — Priority-based Flow Control (PFC)
- 802.1Qaz — Enhanced Transmission Selection (ETS)
- 802.1Qau — Congestion Notification
- Data Center Bridging Exchange (DCBx) protocol

NOTE: Dell Networking OS 9.0(1.2) supports only the PFC, ETS, and DCBx features in data center bridging.

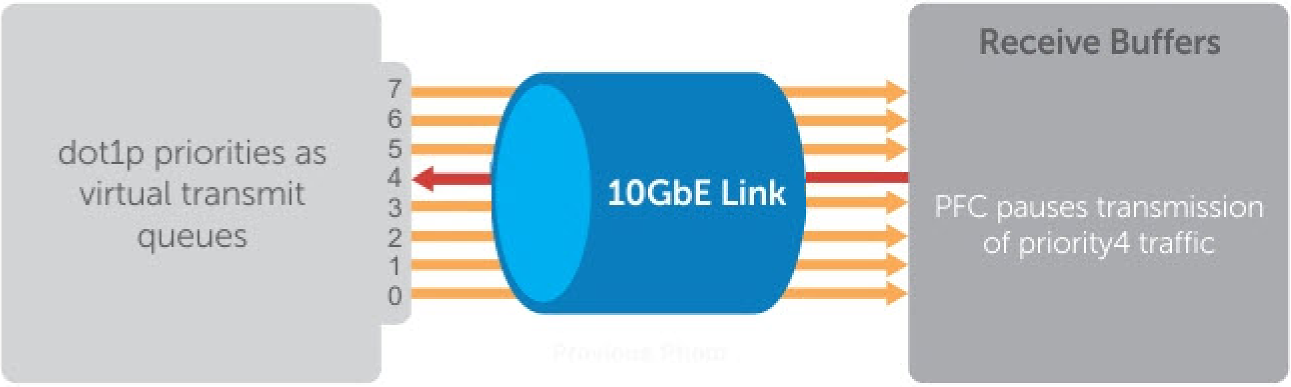
Priority-Based Flow Control

In a data center network, priority-based flow control (PFC) manages large bursts of one traffic type in multiprotocol links so that it does not affect other traffic types and no frames are lost due to congestion.

When PFC detects congestion on a queue for a specified priority, it sends a pause frame for the 802.1p priority traffic to the transmitting device. In this way, PFC ensures that PFC-enabled priority traffic is not dropped by the switch.

PFC enhances the existing 802.3x pause and 802.1p priority capabilities to enable flow control based on 802.1p priorities (classes of service). Instead of stopping all traffic on a link (as performed by the traditional Ethernet pause mechanism), PFC pauses traffic on a link according to the 802.1p priority set on a traffic type. You can create lossless flows for storage and server traffic while allowing for loss in case of LAN traffic congestion on the same physical interface.

The following illustration shows how PFC handles traffic congestion by pausing the transmission of incoming traffic with dot1p priority 4.


dot1p priorities as virtual transmit queues
7
6
5
4
3
2
1
0
10GbE Link
Receive Buffers
PFC pauses transmission of priority4 traffic

The system supports loading two DCB_Config files:

- FCoE converged traffic with priority 3.
- iSCSI storage traffic with priority 4.

In the Dell Networking OS, PFC is implemented as follows:

- PFC is supported on specified 802.1p priority traffic (dot1p 0 to 7) and is configured per interface. However, only two lossless queues are supported on an interface: one for FCoE converged traffic with priority 3 and one for iSCSI storage traffic with priority 4. You must configure the same lossless queues on all ports.
- PFC delay constraints place an upper limit on the transmit time of a queue after receiving a message to pause a specified priority.
- By default, PFC is enabled on an interface with no dot1p priorities configured. You can configure the PFC priorities if the switch negotiates with a remote peer using DCBx
- During DCBx negotiation with a remote peer:
- DCBx communicates with the remote peer by LLDP TLV to determine current policies, such as PFC support and ETS bandwidth allocation.
- If the negotiation fails and PFC is enabled on the port, any user-configured PFC input policies are applied. If no PFC input policy has been previously applied, the PFC default setting is used (no priorities are configured). If you do not enable PFC on an interface, you can enable the 802.3x link-level pause function. By default, the link-level flow pause is disabled when DCBx and PFC are enabled. If no PFC input policy has been applied on the interface, the default PFC settings are used.
- If DCBx negotiation is not successful (for example, a version or TLV mismatch), DCBx is disabled and PFC or ETS cannot be enabled.
- PFC supports buffering to receive data that continues to arrive on an interface while the remote system reacts to the PFC operation.
- PFC uses DCB MIB IEEE 802.1azd2.5 and PFC MIB IEEE 802.1bb-d2.2
- PFC supports buffering to receive data that continues to arrive on an interface while the remote system reacts to the PFC operation.
- PFC uses DCB MIB IEEE 802.1azd2.5 and PFC MIB IEEE 802.1bb-d2.2.
- PFC is supported on specified 802.1p priority traffic (dot1p 0 to 7) and is configured per interface. However, only two lossless queues are supported on an interface: one for Fibre Channel over Ethernet (FCoE) converged traffic and one for Internet Small Computer System Interface (iSCSI) storage traffic. Configure the same lossless queues on all ports.
- PFC supports buffering to receive data that continues to arrive on an interface while the remote system reacts to the PFC operation.

- PFC uses DCB MIB IEEE 802.1azd2.5 and PFC MIB IEEE 802.1bb-d2.2.
- PFC is supported on specified 802.1p priority traffic (dot1p 0 to 7) and is configured per interface. However, only two lossless queues are supported on an interface: one for Fibre Channel over Ethernet (FCoE) converged traffic and one for Internet Small Computer System Interface (iSCSI) storage traffic. Configure the same lossless queues on all ports.
- A dynamic threshold handles intermittent traffic bursts and varies based on the number of PFC priorities contending for buffers, while a static threshold places an upper limit on the transmit time of a queue after receiving a message to pause a specified priority. PFC traffic is paused only after surpassing both static and dynamic thresholds for the priority specified for the port.
- By default, PFC is enabled when you enable DCB. If you have not loaded FCoE_DCB_Config and iSCSI_DCB_Config, DCB is disabled. When you enable DCB globally, you cannot simultaneously enable link-level flow control.
- Buffer space is allocated and de-allocated only when you configure a PFC priority on the port.

Enhanced Transmission Selection

Enhanced transmission selection (ETS) supports optimized bandwidth allocation between traffic types in multiprotocol (Ethernet, FCoE, SCSI) links.

ETS allows you to divide traffic according to its 802.1p priority into different priority groups (traffic classes) and configure bandwidth allocation and queue scheduling for each group to ensure that each traffic type is correctly prioritized and receives its required bandwidth. For example, you can prioritize low-latency storage or server cluster traffic in a traffic class to receive more bandwidth and restrict best-effort LAN traffic assigned to a different traffic class.

Although you can configure strict-priority queue scheduling for a priority group, ETS introduces flexibility that allows the bandwidth allocated to each priority group to be dynamically managed according to the amount of LAN, storage, and server traffic in a flow. Unused bandwidth is dynamically allocated to prioritized priority groups. Traffic is queued according to its 802.1p priority assignment, while flexible bandwidth allocation and the configured queue-scheduling for a priority group is supported.

The following figure shows how ETS allows you to allocate bandwidth when different traffic types are classed according to 802.1p priority and mapped to priority groups.

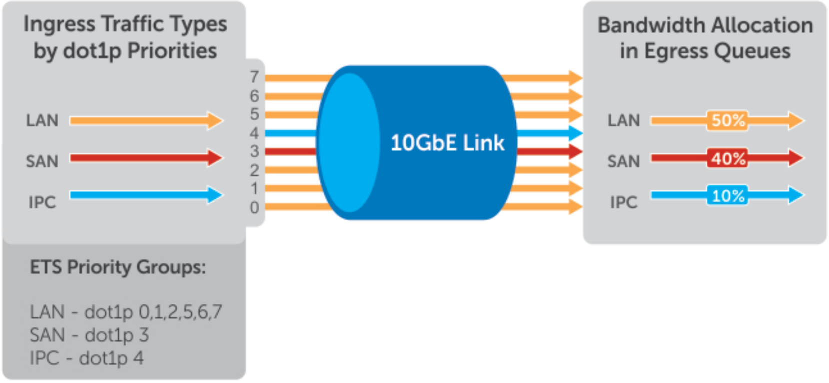


Figure 27. Enhanced Transmission Selection

The following table lists the traffic groupings ETS uses to select multiprotocol traffic for transmission.

Table 10. ETS Traffic Groupings

Traffic Groupings	Description
Priority group	A group of 802.1p priorities used for bandwidth allocation and queue scheduling. All 802.1p priority traffic in a group must have the same traffic handling requirements for latency and frame loss.
Group ID	A 4-bit identifier assigned to each priority group. The range is from 0 to 7 configurable; 8 - 14 reservation and 15.0 - 15.7 is strict priority group..
Group bandwidth	Percentage of available bandwidth allocated to a priority group.
Group transmission selection algorithm (TSA)	Type of queue scheduling a priority group uses.

In Dell Networking OS, ETS is implemented as follows:

- ETS supports groups of 802.1p priorities that have:
  - PFC enabled or disabled
  - No bandwidth limit or no ETS processing
- Bandwidth allocated by the ETS algorithm is made available after strict-priority groups are serviced. Bandwidth is distributed in the following ways:
  - If bandwidth is not assigned to the priority groups, all available bandwidth is equally distributed among the priority groups. For example, if there are two priority groups, each will be assigned 50% of the available bandwidth; if there are four priority groups, each group is assigned 25% of the available bandwidth so that bandwidth use is always 100%.
  - If a priority group does not use its allocated bandwidth, the unused bandwidth is made available to other priority groups so that the sum of the bandwidth use is 100%.
  - If priority group bandwidth is less than 100%, all configured priority group bandwidth is incremented based on the configured percentage ratio until all priority group bandwidth use is 100%.
  - If priority group bandwidth use exceeds 100%, all configured priority group bandwidth is decremented based on the configured percentage ratio until all priority group bandwidth use is 100%.
  - If priority group bandwidth usage is greater than or equal to 100% and any default priority groups exist, then a minimum of 1% bandwidth use is assigned by decreasing 1% of bandwidth from the other priority groups until priority group bandwidth use is 100%.

  NOTE: You must configure at least one ETS priority group to a DCB output policy.
- For ETS traffic selection, an algorithm is applied to priority groups using:
  - Strict priority shaping
  - ETS shaping
  - (Credit-based shaping is not supported.)
- ETS uses the DCB MIB IEEE 802.1azd2.5.

Data Center Bridging Exchange Protocol (DCBx)

The data center bridging exchange (DCBx) protocol is disabled by default on the S4810; ETS is also disabled.

DCBx allows a switch to automatically discover DCB-enabled peers and exchange configuration information. PFC and ETS use DCBx to exchange and negotiate parameters with peer devices. DCBx capabilities include:

- Discovery of DCB capabilities on peer-device connections.
- Determination of possible mismatch in DCB configuration on a peer link.
- Configuration of a peer device over a DCB link.

DCBx requires the link layer discovery protocol (LLDP) to provide the path to exchange DCB parameters with peer devices. Exchanged parameters are sent in organizationally specific TLVs in LLDP data units. For more information, refer to Link Layer Discovery Protocol (LLDP). The following LLDP TLVs are supported for DCB parameter exchange:

PFC parameters	PFC Configuration TLV and Application Priority Configuration TLV.
ETS parameters	ETS Configuration TLV and ETS Recommendation TLV.

Data Center Bridging in a Traffic Flow

The following figure shows how DCB handles a traffic flow on an interface.

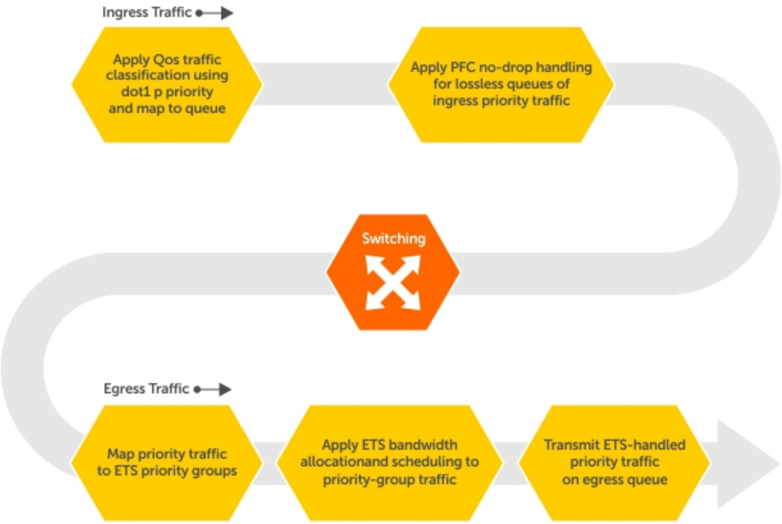

Ingress Traffic
Apply Qos traffic classification using dot1 p priority and map to queue
Apply PFC no-drop handling for lossless queues of ingress priority traffic
Switching
Egress Traffic
Map priority traffic to ETS priority groups
Apply ETS bandwidth allocationand scheduling to priority-group traffic
Transmit ETS-handled priority traffic on egress queue

Figure 28. DCB PFC and ETS Traffic Handling

Enabling Data Center Bridging

DCB is automatically configured when you configure FCoE or iSCSI optimization.
Data center bridging supports converged enhanced Ethernet (CEE) in a data center network. DCB is disabled by default. It must be enabled to support CEE.

- Priority-based flow control
- Enhanced transmission selection
- Data center bridging exchange protocol
- FCoE initialization protocol (FIP) snooping

DCB processes virtual local area network (VLAN)-tagged packets and dot1p priority values. Untagged packets are treated with a dot1p priority of 0.

For DCB to operate effectively, you can classify ingress traffic according to its dot1p priority so that it maps to different data queues. The dot1p-queue assignments used are shown in the following table.

To enable DCB, enable either the iSCSI optimization configuration or the FCoE configuration.

To enable DCB with PFC buffers on a switch, enter the following commands, save the configuration, and reboot the system to allow the changes to take effect.

1. Enable DCB.
   CONFIGURATION mode

   ```
   dcb enable
   ```
2. Set PFC buffering on the DCB stack unit.
   CONFIGURATION mode

   ```
   dcb stack-unit all pfc-buffering pfc-ports 64 pfc-queues 2
   ```

   NOTE: To save the pfc buffering configuration changes, save the configuration and reboot the system.

NOTE: **Dell Networking OS Behavior**: DCB is not supported if you enable link-level flow control on one or more interfaces. For more information, refer to Ethernet Pause Frames.

QoS dot1p Traffic Classification and Queue Assignment

The following section describes QoS dot1P traffic classification and assignments.

DCB supports PFC, ETS, and DCBx to handle converged Ethernet traffic that is assigned to an egress queue according to the following QoS methods:

Honor dot1p	You can honor dot1p priorities in ingress traffic at the port or global switch level (refer to Default dot1p to Queue Mapping) using the `service-class dynamic dot1p` command in INTERFACE configuration mode.
Layer 2 class maps	You can use dot1p priorities to classify traffic in a class map and apply a service policy to an ingress port to map traffic to egress queues.

NOTE: Dell Networking does not recommend mapping all ingress traffic to a single queue when using PFC and ETS. However, Dell Networking does recommend using Ingress traffic classification using the `service-class dynamic dot1p` command (honor dot1p) on all DCB-enabled interfaces. If you use L2 class maps to map dot1p priority traffic to egress queues, take into account the default dot1p-queue assignments in the following table and the maximum number of two lossless queues supported on a port (refer to Configuring Lossless Queues).

Although Dell Networking OS allows you to change the default dot1p priority-queue assignments (refer to Setting dot1p Priorities for Incoming Traffic), DCB policies applied to an interface may become invalid if you reconfigure dot1p-queue mapping. If the configured DCB policy remains valid, the change in the dot1p-queue assignment is allowed.

dot1p Value in the Incoming Frame	Egress Queue Assignment
0	2
1	0
2	1
3	3
4	4
5	5
6	6
7	7

SNMP Support for PFC and Buffer Statistics Tracking

Priority Flow Control (PFC) provides a link level flow control mechanism, which is controlled independently for each frame priority. The goal of this mechanism is to ensure zero loss under congestion in DCB networks.

The SNMP support for monitoring PFC and BST counters and statistics is introduced in Dell Networking OS 9.3(0.1). The enhancement is made on F10-FPSTATS MIB with additional tables to display the PFC and BST counters and statistics.

The following new tables are added in F10-FPSTATS MIB in Dell Networking OS 9.3(0.1):

- fpEgrQBuffSnapshotTable
- fpIngPgBuffSnapshotTable
- fpStatsPerPgTable
- pfcPerPrioTable

fpEgrQBuffSnapshotTable	This table fetches the BST statistics at Egress Port with respect to the buffer used. This table displays the Snapshot of the Buffer cells used by Unicast and Multicast Data and Control Queues.
fpIngPgBuffSnapshotTable	This table fetches the BST statistics at the Ingress Port with respect to the Shared Cells and the Headroom cells used per Priority Group. The snapshot of the Ingress Shared cells used and the Ingress Headroom cells used per Priority Group, when PFC is enabled, will be displayed in this table. This table is indexed by stack-unit index, port number and the priority group number.

fpStatsPerPgTable	This table fetches the Allocated Min cells, Shared cells and Headroom cells per Priority Group, the mode in which the buffer cells are allocated - Static or Dynamic and the Used Min Cells, Shared cells and Headroom cells per Priority Group. The table fetches a value of 0 if the mode of allocation is Static and a value of 1 if the mode of allocation is Dynamic. This table is indexed by stack-unit number, port number and priority group number.
pfcPerPrioTable	This table fetches the number of PFC frames transmitted (PFC Requests) and the number of PFC frames received (PFC Indications) per priority on a per port basis. This table is indexed by the stack-unit index, port number and priority.

Configuring Priority-Based Flow Control

PFC provides a flow control mechanism based on the 802.1p priorities in converged Ethernet traffic received on an interface and is enabled by default when you enable DCB.
As an enhancement to the existing Ethernet pause mechanism, PFC stops traffic transmission for specified priorities (Class of Service (CoS) values) without impacting other priority classes. Different traffic types are assigned to different priority classes.

When traffic congestion occurs, PFC sends a pause frame to a peer device with the CoS priority values of the traffic that is to be stopped. Data Center Bridging Exchange protocol (DCBx) provides the link-level exchange of PFC parameters between peer devices. PFC allows network administrators to create zero-loss links for Storage Area Network (SAN) traffic that requires no-drop service, while retaining packet-drop congestion management for Local Area Network (LAN) traffic.

To ensure complete no-drop service, apply the same DCB input policy with the same pause time and dot1p priorities on all PFC-enabled peer interfaces.

To configure PFC and apply a PFC input policy to an interface, follow these steps.

1. Create a DCB input policy to apply pause or flow control for specified priorities using a configured delay time.

   CONFIGURATION mode

   ```
   dcb-input policy-name
   ```

   The maximum is 32 alphanumeric characters.
2. Configure the link delay used to pause specified priority traffic.

   DCB INPUT POLICY mode

   ```
   pfc link-delay value
   ```

   One quantum is equal to a 512-bit transmission.

   The range (in quanta) is from 712 to 65535.

   The default is **45556** quantum in link delay.
3. Configure the CoS traffic to be stopped for the specified delay.

   DCB INPUT POLICY mode

   ```
   pfc priority priority-range
   ```

Enter the 802.1p values of the frames to be paused.

The range is from 0 to 7.

The default is **none**.

Maximum number of loss less queues supported on the switch: 2.

Separate priority values with a comma. Specify a priority range with a dash, for example: pfc priority 1,3,5-7.

4. Enable the PFC configuration on the port so that the priorities are included in DCBx negotiation with peer PFC devices.
   DCB INPUT POLICY mode

   ```
   pfc mode on
   ```

   The default is PFC mode is **on**.
5. (Optional) Enter a text description of the input policy.
   DCB INPUT POLICY mode

   ```
   description text
   ```

   The maximum is 32 characters.
6. Exit DCB input policy configuration mode.
   DCB INPUT POLICY mode

   ```
   exit
   ```
7. Enter interface configuration mode.
   CONFIGURATION mode

   ```
   interface type slot/port
   ```
8. Apply the input policy with the PFC configuration to an ingress interface.
   INTERFACE mode

   ```
   dcb-policy input policy-name
   ```
9. Repeat Steps 1 to 8 on all PFC-enabled peer interfaces to ensure lossless traffic service.

Dell Networking OS Behavior: As soon as you apply a DCB policy with PFC enabled on an interface, DCBx starts exchanging information with PFC-enabled peers. The IEEE802.1Qbb, CEE, and CIN versions of PFC Type, Length, Value (TLV) are supported. DCBx also validates PFC configurations that are received in TLVs from peer devices.

By applying a DCB input policy with PFC enabled, you enable PFC operation on ingress port traffic. To achieve complete lossless handling of traffic, also enable PFC on all DCB egress ports or configure the dot1p priority-queue assignment of PFC priorities to lossless queues (refer to Configuring Lossless Queues).

To remove a DCB input policy, including the PFC configuration it contains, use the `no dcb-input policy-name` command in INTERFACE Configuration mode. To disable PFC operation on an interface, use the `no pfc mode on` command in DCB Input Policy Configuration mode. PFC is enabled and disabled as the global DCB operation is enabled (`dcb enable`) or disabled (`no dcb enable`).

You can enable any number of 802.1p priorities for PFC. Queues to which PFC priority traffic is mapped are lossless by default. Traffic may be interrupted due to an interface flap (going down and coming up) when you reconfigure the lossless queues for no-drop priorities in a PFC input policy and reapply the policy to an interface.

To apply PFC, a PFC peer must support the configured priority traffic (as detected by DCBx).

To honor a PFC pause frame multiplied by the number of PFC-enabled ingress ports, the minimum link delay must be greater than the round-trip transmission time the peer requires.

If you apply an input policy with PFC disabled (`no pfc mode on`):

- You can enable link-level flow control on the interface (refer to Ethernet Pause Frames). To delete the input policy, first disable link-level flow control. PFC is then automatically enabled on the interface because an interface is by default PFC-enabled.
- PFC still allows you to configure lossless queues on a port to ensure no-drop handling of lossless traffic (refer to Configuring Lossless Queues).

NOTE: You cannot enable PFC and link-level flow control at the same time on an interface.

When you apply an input policy to an interface, an error message displays if:

- The PFC dot1p priorities result in more than two lossless port queues globally on the switch.
- Link-level flow control is already enabled. You cannot be enable PFC and link-level flow control at the same time on an interface.
- In a switch stack, configure all stacked ports with the same PFC configuration.

A DCB input policy for PFC applied to an interface may become invalid if you reconfigure dot1p-queue mapping (refer to the *Create Input Policy Maps* section in the Quality of Service (QoS) chapter). This situation occurs when the new dot1p-queue assignment exceeds the maximum number (2) of lossless queues supported globally on the switch. In this case, all PFC configurations received from PFC-enabled peers are removed and re-synchronized with the peer devices.

Traffic may be interrupted when you reconfigure PFC no-drop priorities in an input policy or reapply the policy to an interface.

Dell Networking OS does not support MACsec Bypass Capability (MBC).

Configuring Lossless Queues

DCB also supports the manual configuration of lossless queues on an interface when PFC mode is turned off and priority classes are disabled in a DCB input policy applied to the interface.
Prerequisite: A DCB input policy with PFC configuration is applied to the interface with the following conditions:

- PFC mode is off (no pfc mode on).
- No PFC priority classes are configured (`no pfc priority priority-range`).

The configuration of no-drop queues provides flexibility for ports on which PFC is not needed but lossless traffic should egress from the interface.

Lossless traffic egresses out the no-drop queues. Ingress dot1p traffic from PFC-enabled interfaces is automatically mapped to the no-drop egress queues.

1. Enter INTERFACE Configuration mode.

   CONFIGURATION mode

```
interface type slot/port
```

2. Configure the port queues that will still function as no-drop queues for lossless traffic.

   INTERFACE mode

   ```
   pfc no-drop queues queue-range
   ```

   For the dot1p-queue assignments, refer to the dot1p Priority-Queue Assignment table.

   The maximum number of lossless queues globally supported on the switch is two.

   The range is from 0 to 7. Separate the queue values with a comma; specify a priority range with a dash; for example, `pfc no-drop queues 1,7` or `pfc no-drop queues 2-7`.

   The range is from 0 to 3. Separate the queue values with a comma; specify a priority range with a dash; for example, `pfc no-drop queues 1,3` or `pfc no-drop queues 2-3`.

   The default: No lossless queues are configured.

NOTE: **Dell Networking OS Behavior**: By default, no lossless queues are configured on a port.

A limit of two lossless queues is supported on a port. If the amount of priority traffic that you configure to be paused exceeds the two lossless queues, an error message displays. Reconfigure the input policy using a smaller number of PFC priorities.

If you configure lossless queues on an interface that already has a DCB input policy with PFC enabled (**pfc mode on**), an error message displays.

Traffic may be interrupted due to an interface flap (going down and coming up) when you reconfigure lossless queues on no-drop priorities in an input policy and re-apply the policy to an interface.

In S6000, any pfc-dot1p priorities configured on a given interface need not be the same across the system. In other words, lossless queue limit is applicable on a per-port level and not on the global-config context. For example, one of the Te/Fo interfaces can have pfc-dot1p priorities as 2 and 3. Whereas, the other Te/Fo interface(s) can have its pfc-dot1p priorities as 4 and 5.

It is the user responsibility to have symmetric PFC configurations on the interfaces involved in a particular PFC-enabled traffic-flow to obtain lossless behavior

Configuring the PFC Buffer in a Switch Stack

In a switch stack, you must configure all stacked ports with the same PFC configuration. In addition, you must configure a separate buffer of memory allocated exclusively to a service pool accessed by queues on which priority-based control flows are mapped.

These PFC-enabled queues ensure the lossless transmission of storage and server traffic. The buffer required for the PFC service pool is calculated based on the number of ports and port queues used by PFC traffic.

You can configure the size of the PFC buffer for all switches in a stack or all port pipes on a specified stack unit by entering the following commands on the master switch.

- Configure the PFC buffer for all switches in the stack.

  CONFIGURATION mode

  ```
  [no] dcb stack-unit all pfc-buffering pfc-port {1-64} pfc-queues {1-2}
  ```

By default, the PFC buffer is enabled on all ports on the stack unit.

- Configure the PFC buffer for all port pipes in a specified stack unit by specifying the port-pipe number, number of PFC-enabled ports, and number of configured lossless queues.

  CONFIGURATION mode

  ```
  [no] dcb stack-unit stack-unit-id [port-set port-set-id] pfc-buffering pfc-
  ports {1-64} pfc-queues {1-2}
  ```

  Valid stack-unit IDs are 0 to 5.

  The only valid port-set ID (port-pipe number) is 0.

Dell Networking OS Behavior: If you configure PFC on a 40GbE port, count the 40GbE port as four PFC-enabled ports in the pfc-port number you enter in the command syntax.

To achieve lossless PFC operation, the PFC port count and queue number used for the reserved buffer size that is created must be greater than or equal to the buffer size required for PFC-enabled ports and lossless queues on the switch.

For the PFC buffer configuration to take effect, you must reload the stack or a specified stack unit (use the `reload` command at EXEC Privilege level).

Configure Enhanced Transmission Selection

ETS provides a way to optimize bandwidth allocation to outbound 802.1p classes of converged Ethernet traffic.

Different traffic types have different service needs. Using ETS, you can create groups within an 802.1p priority class to configure different treatment for traffic with different bandwidth, latency, and best-effort needs.

For example, storage traffic is sensitive to frame loss; interprocess communication (IPC) traffic is latency-sensitive. ETS allows different traffic types to coexist without interruption in the same converged link by:

- Allocating a guaranteed share of bandwidth to each priority group.
- Allowing each group to exceed its minimum guaranteed bandwidth if another group is not fully using its allotted bandwidth.

To configure ETS and apply an ETS output policy to an interface, you must:

1. Create a Quality of Service (QoS) output policy with ETS scheduling and bandwidth allocation settings.
2. Create a priority group of 802.1p traffic classes.
3. Configure a DCB output policy in which you associate a priority group with a QoS DCB output policy.
4. Apply the DCB output policy to an interface.

ETS Prerequisites and Restrictions

The following prerequisites and restrictions apply when you configure ETS bandwidth allocation or queue scheduling and apply a DCB output policy on an interface.

- Configuring ETS bandwidth allocation or a queue scheduler for dot1p priorities in a priority group is applicable if the DCBx version used on a port is CIN (refer to Configuring DCBx).

- When allocating bandwidth or configuring a queue scheduler for dot1p priorities in a priority group on a DCBx CIN interface, take into account the CIN bandwidth allocation (refer to Configuring Bandwidth Allocation for DCBx CIN) and dot1p-queue mapping.
- Although an DCB Output policy does not support WRED, ECN, rate shaping, and rate limiting because DCBx does not negotiate these parameters with peer devices, you can apply a QoS output policy with WRED and/or rate shaping on a DCBx CIN-enabled interface (refer to Configuring Port-Based Rate Shaping and Weighted Random Early Detection). In this case, the WRED or rate shaping configuration in the QoS output policy takes into account the bandwidth allocation or queue scheduler configured in the DCB Output policy.
- You can only use a QoS DCB output policy in association with a priority group in a DCB output policy and cannot be applied to an interface as a normal QoS output policy (refer to Applying an ETS Output Policy for a Priority Group to an Interface and Creating an Output QoS Policy in the Quality of Service (QoS) chapter.).

NOTE: The IEEE 802.1Qaz, CEE, and CIN versions of ETS are supported.

Creating a QoS DCB Output Policy

A QoS output policy that you create to optimize bandwidth on an output interface for specified priority traffic consists of the ETS settings (the bandwidth percentage and queue schedule) used in DCBx negotiations with peer devices.

1. Create a QoS output policy to configure the ETS bandwidth allocation and scheduling for priority traffic.

   CONFIGURATION mode

   ```
   qos-policy-output policy-name ets
   ```

   The maximum is 32 characters.
2. (Optional) Configure the method used to schedule priority traffic in port queues.

   POLICY-MAP-OUT-ETS mode

   ```
   scheduler value
   ```

   strict — Strict priority traffic is serviced before any other queued traffic (refer to Enabling Strict-Priority Queueing in the Quality of Service (QoS) chapter).

   NOTE: If you configure a scheduling method, you cannot configure bandwidth allocation in Step 3.

   werr —Weighted elastic round robin provides low-latency scheduling for priority traffic on port queues. WERR scheduling is used to queue priority traffic by default.
3. (Optional) Configure the bandwidth percentage allocated to priority traffic in port queues.

   POLICY-MAP-OUT-ETS mode

   ```
   bandwidth-percentage percentage
   ```

   The percentage range is from 1 to 100% in units of 1%. The sum of bandwidth percentage assigned to dot1p priorities/queues in a priority group should be 100%.

   The default is **none**.

   NOTE: If you configure bandwidth allocation, you cannot configure a scheduling method in Step 2.

4. Exit DCB Output Policy Configuration mode.
   POLICY-MAP-OUT-ETS mode

   ```
   exit
   ```

Dell Networking OS Behavior: Traffic in priority groups is assigned to strict-queue or WERR scheduling in an DCB output policy and is managed using the ETS bandwidth-assignment algorithm. Dell Networking OS deqeues all frames of strict-priority traffic before servicing any other queues. A queue with strict-priority traffic can starve other queues in the same port.

ETS-assigned bandwidth allocation and scheduling apply only to data queues, not to control queues.

Dell Networking OS supports hierarchical scheduling on an interface. Dell Networking OS control traffic is redirected to control queues as higher priority traffic with strict-priority scheduling. After control queues drain out, the remaining data traffic is scheduled to queues according to the bandwidth and scheduler configuration in the DCB output policy. The available bandwidth (that the ETS algorithm calculates) is equal to the link bandwidth after scheduling non-ETS higher-priority traffic.

The configuration of bandwidth allocation and strict-queue scheduling is not supported at the same time for a priority group. If both are configured, the configured bandwidth allocation is ignored for priority-group traffic when you apply the output policy on an interface (refer to Applying an ETS Output Policy for a Priority Group to an Interface).

Bandwidth assignment in a dot.1p priority-queue: By default, equal bandwidth is assigned to each port queue and each dot1p priority in a priority group. To configure bandwidth amounts in associated dot1p queues, use the `bandwidth-percentage` command. When specified bandwidth is assigned to some port queues and not to others, the remaining bandwidth (100% minus assigned bandwidth amount) is equally distributed to unassigned nonstrict priority queues in the priority group. The sum of the allocated bandwidth to all queues in a priority group should be 100% of the bandwidth on the link.

Bandwidth assignment in a priority group: By default, equal bandwidth is assigned to each priority group in the DCB output policy applied to an egress port if you did not configure bandwidth allocation. The sum of configured bandwidth allocation to dot1p priority traffic in all ETS priority groups must be 100%. Allocate at least 1% of the total bandwidth to each priority group and queue. If you assign bandwidth to some priority groups but not to others, the remaining bandwidth (100% minus assigned bandwidth amount) is equally distributed to non strict-priority groups which have no configured scheduler.

Scheduling of priority traffic: dot1p priority traffic on the switch is scheduled to the current queue mapping. dot1p priorities within the same queue should have the same traffic properties and scheduling method.

DCB output-policy error: If an error occurs in an DCB output-policy configuration, the configuration is ignored and the scheduler and bandwidth allocation settings are reset to the ETS default values (all priorities are in the same ETS priority group and bandwidth is allocated equally to each priority). If an error occurs when a port receives a peer's ETS configuration, the port's configuration is reset to the previously configured DCB output policy. If no DCB output policy was previously applied, the port is reset to the default ETS parameters.

Creating an ETS Priority Group

An ETS priority group specifies the range of 802.1p priority traffic to which a QoS output policy with ETS settings is applied on an egress interface. You can associate a priority group to more than one ETS output policy on different interfaces.

1. Create an ETS priority group to use with an ETS output policy.

   CONFIGURATION mode

   ```
   priority-group group-name
   ```

   The maximum is 32 characters.
2. Configure the priority-group identifier.

   PRIORITY-GROUP mode

   ```
   set-pgid value
   ```

   The range is from 0 to 7.

   The default is **none**.
3. Configure the 802.1p priorities for the traffic on which you want to apply an ETS output policy.

   PRIORITY-GROUP mode

   ```
   priority-list value
   ```

   The range is from 0 to 7.

   The default is **none**.

   Separate priority values with a comma. Specify a priority range with a dash. For example, priority-list 3,5-7.
4. Exit priority-group configuration mode.

   PRIORITY-GROUP mode

   ```
   exit
   ```
5. Repeat Steps 1 to 4 to configure all remaining dot1p priorities in an ETS priority group.

Dell Networking OS Behavior: A priority group consists of 802.1p priority values that are grouped for similar bandwidth allocation and scheduling, and that share latency and loss requirements. All 802.1p priorities mapped to the same queue must be in the same priority group.

Configure all 802.1p priorities in priority groups associated with an ETS output policy (refer to Applying an ETS Output Policy for a Priority Group to an Interface). You can assign each dot1p priority to only one priority group.

By default, all 802.1p priorities are grouped in priority group 0 and 100% of the port bandwidth is assigned to priority group 0. The complete bandwidth is equally assigned to each priority class so that each class has 12 to 13%.

The maximum number of priority groups supported in ETS output policies on an interface is equal to the number of data queues (4) on the port. The 802.1p priorities in a priority group can map to multiple queues.

If you configure more than one priority queue as strict priority or more than one priority group as strict priority, the higher numbered priority queue is given preference when scheduling data traffic.

Applying an ETS Output Policy for a Priority Group to an Interface

To apply ETS on egress port traffic, you must associate a priority group with an ETS output policy which has scheduling and bandwidth configuration in a DCB output policy, and then apply the output policy to an interface.

1. Create a DCB output policy to associate an ETS configuration with priority traffic.
   CONFIGURATION mode

   ```
   dcb-output policy-name
   ```

   The maximum is 32 alphanumeric characters.
2. Enable the ETS configuration so that scheduling and bandwidth allocation configured in an ETS output policy or received in a DCBx TLV from a peer can take effect on an interface.
   DCB OUTPUT POLICY mode

   ```
   ets mode on
   ```

   The default: ETS mode is on.
3. Associate the 802.1p priority traffic in a priority group with the ETS configuration in a QoS output policy.
   DCB OUTPUT POLICY mode

   ```
   priority-group group-name qos-policy ets-policy-name
   ```

4. (Optional) Enter a text description of the output policy.
   DCB OUTPUT POLICY mode

   ```
   description text
   ```

   The maximum is 32 characters.
5. Repeat Steps 1 to 4 to configure all remaining ETS priority groups with an ETS output policy.
6. Exit DCB Output Policy Configuration mode.
   DCB OUTPUT POLICY

   ```
   exit
   ```
7. Enter INTERFACE Configuration mode.
   CONFIGURATION mode

   ```
   interface type slot/port
   ```
8. Apply the output policy with the ETS configuration to an egress interface.
   INTERFACE mode

   ```
   dcb-policy output policy-name
   ```

Dell Networking OS Behavior: Create a DCB output policy to associate a priority group with an ETS output policy with scheduling and bandwidth configuration. You can apply a DCB output policy on multiple egress ports.

The ETS configuration associated with 802.1p priority traffic in a DCB output policy is used in DCBx negotiation with ETS peers.

When you apply an ETS output policy to an interface, ETS-configured scheduling and bandwidth allocation take precedence over any configured settings in the QoS output policies.

To remove an ETS output policy from an interface, use the `no dcb-policy output policy-name` command. DCB and ETS are both disabled by default. When DCB is enabled, ETS is enabled on all interfaces that have the default ETS configuration applied.

If you disable ETS in an output policy applied to an interface (the `no ets mode on` command), any previously configured QoS settings at the interface or global level take effect. If QoS settings are configured at the interface or global level and in an output policy map (the `service-policy output` command), the QoS configuration in the output policy take precedence.

ETS Operation with DCBx

The following section describes DCBx negotiation with peer ETS devices.

In DCBx negotiation with peer ETS devices, ETS configuration is handled as follows:

- ETS TLVs are supported in DCBx versions CIN, CEE, and IEEE2.5.
- The DCBx port-role configurations determine the ETS operational parameters (refer to Configure a DCBx Operation).
- ETS configurations received from TLVs from a peer are validated.
- If there is a hardware limitation or TLV error:
  - DCBx operation on an ETS port goes down.
  - New ETS configurations are ignored and existing ETS configurations are reset to the previously configured ETS output policy on the port or to the default ETS settings if no ETS output policy was previously applied.
- ETS operates with legacy DCBx versions as follows:
  - In the CEE version, the priority group/traffic class group (TCG) ID 15 represents a non-ETS priority group. Any priority group configured with a scheduler type is treated as a strict-priority group and is given the priority-group (TCG) ID 15.
  - The CIN version supports two types of strict-priority scheduling:
    - Group strict priority: Use this to increase its bandwidth usage to the bandwidth total of the priority group and allow a single priority flow in a priority group. A single flow in a group can use all the bandwidth allocated to the group.
    - Link strict priority: Use this to increase to the maximum link bandwidth and allow a flow in any priority group.

  CIN supports only the dot1p priority-queue assignment in a priority group. To configure a dot1p priority flow in a priority group to operate with link strict priority, you configure: The dot1p priority for strict-priority scheduling (`strict-priority` command). The priority group for strict-priority scheduling (`scheduler strict` command; Creating a QoS ETS Output Policy).

  If you configure only the priority group in an ETS output policy or only the dot1p priority for strict-priority scheduling, the flow is handled with group strict priority.

Configuring Bandwidth Allocation for DCBx CIN

After you apply an ETS output policy to an interface, if the DCBx version used in your data center network is CIN, you may need to configure a QoS output policy to overwrite the default CIN bandwidth allocation. This default setting divides the bandwidth allocated to each port queue equally between the dot1p priority traffic assigned to the queue.

To create a QoS output policy that allocates different amounts of bandwidth to the different traffic types/ dot1p priorities assigned to a queue and apply the output policy to the interface, follow these steps.

1. Create a QoS output policy.

   CONFIGURATION mode

   ```
   qos-policy-output output-policy-name
   ```

   The maximum 32 alphanumeric characters.
2. Configure the percentage of bandwidth to allocate to the dot1p priority/queue traffic in the associated L2 class map.

   QoS OUTPUT POLICY mode

   ```
   bandwidth-percentage percentage
   ```

   The default is **none**.
3. Repeat Step 2 to configure bandwidth percentages for other priority queues on the port.

   QoS OUTPUT POLICY mode

   ```
   bandwidth-percentage percentage
   ```
4. Exit QoS Output Policy Configuration mode.

   QoS OUTPUT POLICY mode

   ```
   exit
   ```
5. Enter INTERFACE Configuration mode.

   CONFIGURATION mode

   ```
   interface type slot/port
   ```
6. Apply the QoS output policy with the bandwidth percentage for specified priority queues to an egress interface.

   INTERFACE mode

   ```
   service-policy output output-policy-name
   ```

Applying DCB Policies in a Switch Stack

You can apply a DCB input policy with PFC configuration to all stacked ports in a switch stack or on a stacked switch. You can apply different DCB input policies to different stacked switches.
To apply DCB policies in a switch stack, use the following command.

- Apply the specified DCB input policy on all ports of the switch stack or a single stacked switch.

  CONFIGURATION mode

```
dcb-policy input stack-unit {all | stack-unit-id} stack-ports all dcb-input-
policy-name
```

Entering this command removes all DCB input policies applied to stacked ports.

A `dcb-policy input stack-unit all` command overwrites any previous `dcb-policy input stack-unit` *`stack-unit-id`* configurations. Similarly, a `dcb-policy input stack-unit` *`stack-unit-id`* command overwrites any previous `dcb-policy input stack-unit all` configuration.

Entering the `no dcb-policy input stack-unit all` command removes all DCB input policies applied to stacked ports and resets PFC to its default settings. The `no dcb-policy input stack-unit` *`stack-unit-id`* command removes only the DCB input policy applied to the specified switch.

Applying DCB Policies with an ETS Configuration

You can apply a DCB output policy with ETS configuration to all stacked ports in a switch stack or an individual stacked switch. In addition, you can apply different DCB output policies to different stack units.

- Apply the specified DCB output policy on all ports of the switch stack or a stacked switch.
  CONFIGURATION mode

  ```
  dcb-policy output stack-unit {all | stack-unit-id} stack-ports all dcb-
  output-policy-name
  ```

  Entering this command removes all DCB input policies applied to stacked ports.

A `dcb-policy output stack-unit all` command overwrites any previous `dcb-policy output stack-unit` *`stack-unit-id`* configurations. Similarly, a `dcb-policy output stack-unit` *`stack-unit-id`* command overwrites any previous `dcb-policy output stack-unit all` configuration.

Entering the `no dcb-policy output stack-unit all` command removes all DCB output policies applied to stacked ports. The `no dcb-policy output stack-unit` *`stack-unit-id`* command removes only the DCB output policy applied to the specified switch.

Configure a DCBx Operation

DCB devices use data center bridging exchange protocol (DCBx) to exchange configuration information with directly connected peers using the link layer discovery protocol (LLDP) protocol.

DCBx can detect the misconfiguration of a peer DCB device, and optionally, configure peer DCB devices with DCB feature settings to ensure consistent operation in a data center network.

DCBx is a prerequisite for using DCB features, such as priority-based flow control (PFC) and enhanced traffic selection (ETS), to exchange link-level configurations in a converged Ethernet environment. DCBx is also deployed in topologies that support lossless operation for FCoE or iSCSI traffic. In these scenarios, all network devices are DCBx-enabled (DCBx is enabled end-to-end). For more information about how these features are implemented and used, refer to:

- Configure Enhanced Transmission Selection

DCBx supports the following versions: CIN, CEE, and IEEE2.5.

Prerequisite: For DCBx, enable LLDP on all DCB devices.

DCBx Operation

DCBx performs the following operations:

- Discovers DCB configuration (such as PFC and ETS) in a peer device.
- Detects DCB mis-configuration in a peer device; that is, when DCB features are not compatibly configured on a peer device and the local switch. Mis-configuration detection is feature-specific because some DCB features support asymmetric configuration.
- Reconfigures a peer device with the DCB configuration from its configuration source if the peer device is willing to accept configuration.
- Accepts the DCB configuration from a peer if a DCBx port is in “willing” mode to accept a peer’s DCB settings and then internally propagates the received DCB configuration to its peer ports.

DCBx Port Roles

To enable the auto-configuration of DCBx-enabled ports and propagate DCB configurations learned from peer DCBx devices internally to other switch ports, use the following DCBx port roles.

Auto-upstream

The port advertises its own configuration to DCBx peers and is *willing* to receive peer configuration. The port also propagates its configuration to other ports on the switch.

The first auto-upstream that is capable of receiving a peer configuration is elected as the configuration source. The elected configuration source then internally propagates the configuration to other auto-upstream and auto-downstream ports. A port that receives an internally propagated configuration overwrites its local configuration with the new parameter values. When an auto-upstream port (besides the configuration source) receives and overwrites its configuration with internally propagated information, one of the following actions is taken:

- If the peer configuration received is compatible with the internally propagated port configuration, the link with the DCBx peer is enabled.
- If the received peer configuration is not compatible with the currently configured port configuration, the link with the DCBx peer port is disabled and a syslog message for an incompatible configuration is generated. The network administrator must then reconfigure the peer device so that it advertises a compatible DCB configuration.
  - The configuration received from a DCBx peer or from an internally propagated configuration is not stored in the switch’s running configuration.
  - On a DCBx port in an auto-upstream role, the PFC and application priority TLVs are enabled. ETS recommend TLVs are disabled and ETS configuration TLVs are enabled.

Auto-downstream

The port advertises its own configuration to DCBx peers but is *not willing* to receive remote peer configuration. The port always accepts internally propagated configurations from a configuration source. An auto-downstream port that receives an internally propagated configuration overwrites its local configuration with the new parameter values.

When an auto-downstream port receives and overwrites its configuration with internally propagated information, one of the following actions is taken:

- If the peer configuration received is compatible with the internally propagated port configuration, the link with the DCBx peer is enabled.

- If the received peer configuration is not compatible with the currently configured port configuration, the link with the DCBx peer port is disabled and a syslog message for an incompatible configuration is generated. The network administrator must then reconfigure the peer device so that it advertises a compatible DCB configuration.
    - The internally propagated configuration is not stored in the switch's running configuration.
    - On a DCBx port in an auto-downstream role, all PFC, application priority, ETS recommend, and ETS configuration TLVs are enabled.

Configuration source	The port is configured to serve as a source of configuration information on the switch. Peer DCB configurations received on the port are propagated to other DCBx auto-configured ports. If the peer configuration is compatible with a port configuration, DCBx is enabled on the port. On a configuration-source port, the link with a DCBx peer is enabled when the port receives a DCB configuration that can be internally propagated to other auto-configured ports. The configuration received from a DCBx peer is not stored in the switch’s running configuration. On a DCBx port that is the configuration source, all PFC and application priority TLVs are enabled. ETS recommend TLVs are disabled and ETS configuration TLVs are enabled.
Manual	The port is configured to operate only with administrator-configured settings and does not auto-configure with DCB settings received from a DCBx peer or from an internally propagated configuration from the configuration source. If you enable DCBx, ports in Manual mode advertise their configurations to peer devices but do not accept or propagate internal or external configurations. Unlike other user-configured ports, the configuration of DCBx ports in Manual mode is saved in the running configuration. On a DCBx port in a manual role, all PFC, application priority, ETS recommend, and ETS configuration TLVs are enabled. When making a configuration change to a DCBx port in a Manual role, Dell Networking recommends shutting down the interface using the `shutdown` command, change the configuration, then re-activate the interface using the `no shutdown` command.

The default for the DCBx port role is **manual**.

NOTE: On a DCBx port, application priority TLV advertisements are handled as follows:

- The application priority TLV is transmitted only if the priorities in the advertisement match the configured PFC priorities on the port.
- On auto-upstream and auto-downstream ports:
  - If a configuration source is elected, the ports send an application priority TLV based on the application priority TLV received on the configuration-source port. When an application priority TLV is received on the configuration-source port, the auto-upstream and auto-downstream ports use the internally propagated PFC priorities to match against the received application priority. Otherwise, these ports use their locally configured PFC priorities in application priority TLVs.
  - If no configuration source is configured, auto-upstream and auto-downstream ports check to see that the locally configured PFC priorities match the priorities in a received application priority TLV.
- On manual ports, an application priority TLV is advertised only if the priorities in the TLV match the PFC priorities configured on the port.

DCB Configuration Exchange

The DCBx protocol supports the exchange and propagation of configuration information for the enhanced transmission selection (ETS) and priority-based flow control (PFC) DCB features.

DCBx uses the following methods to exchange DCB configuration parameters:

Asymmetric	DCB parameters are exchanged between a DCBx-enabled port and a peer port without requiring that a peer port and the local port use the same configured values for the configurations to be compatible. For example, ETS uses an asymmetric exchange of parameters between DCBx peers.
Symmetric	DCB parameters are exchanged between a DCBx-enabled port and a peer port but requires that each configured parameter value be the same for the configurations in order to be compatible. For example, PFC uses an symmetric exchange of parameters between DCBx peers.

Configuration Source Election

When an auto-upstream or auto-downstream port receives a DCB configuration from a peer, the port first checks to see if there is an active configuration source on the switch.

- If a configuration source already exists, the received peer configuration is checked against the local port configuration. If the received configuration is compatible, the DCBx marks the port as DCBx-enabled. If the configuration received from the peer is not compatible, a warning message is logged and the DCBx frame error counter is incremented. Although DCBx is operationally disabled, the port keeps the peer link up and continues to exchange DCBx packets. If a compatible peer configuration is later received, DCBx is enabled on the port.
- If there is no configuration source, a port may elect itself as the configuration source. A port may become the configuration source if the following conditions exist:
  - No other port is the configuration source.
  - The port role is auto-upstream.
  - The port is enabled with link up and DCBx enabled.
  - The port has performed a DCBx exchange with a DCBx peer.
  - The switch is capable of supporting the received DCB configuration values through either a symmetric or asymmetric parameter exchange.

A newly elected configuration source propagates configuration changes received from a peer to the other auto-configuration ports. Ports receiving auto-configuration information from the configuration source ignore their current settings and use the configuration source information.

Propagation of DCB Information

When an auto-upstream or auto-downstream port receives a DCB configuration from a peer, the port acts as a DCBx client and checks if a DCBx configuration source exists on the switch.

- If a configuration source is found, the received configuration is checked against the currently configured values that are internally propagated by the configuration source. If the local configuration is compatible with the received configuration, the port is enabled for DCBx operation and synchronization.
- If the configuration received from the peer is not compatible with the internally propagated configuration used by the configuration source, the port is disabled as a client for DCBx operation and synchronization and a syslog error message is generated. The port keeps the peer link up and continues to exchange DCBx packets. If a compatible configuration is later received from the peer, the port is enabled for DCBx operation.

NOTE: DCB configurations internally propagated from a configuration source do not overwrite the configuration on a DCBx port in a manual role. When a configuration source is elected, all auto-upstream ports other than the configuration source are marked as *willing disabled*. The internally propagated DCB configuration is refreshed on all auto-configuration ports and each port may begin configuration negotiation with a DCBx peer again.

Auto-Detection and Manual Configuration of the DCBx Version

When operating in Auto-Detection mode (the `DCBx version auto` command), a DCBx port automatically detects the DCBx version on a peer port. Legacy CIN and CEE versions are supported in addition to the standard IEEE version 2.5 DCBx.

A DCBx port detects a peer version after receiving a valid frame for that version. The local DCBx port reconfigures to operate with the peer version and maintains the peer version on the link until one of the following conditions occurs:

- The switch reboots.
- The link is reset (goes down and up).
- User-configured CLI commands require the version negotiation to restart.
- The peer times out.
- Multiple peers are detected on the link.

If you configure a DCBx port to operate with a specific version (the `DCBx version {cee | cin | ieee-v2.5}` command in the Configuring DCBx), DCBx operations are performed according to the configured version, including fast and slow transmit timers and message formats. If a DCBx frame with a different version is received, a syslog message is generated and the peer version is recorded in the peer status table. If the frame cannot be processed, it is discarded and the discard counter is incremented.

NOTE: Because DCBx TLV processing is best effort, it is possible that CIN frames may be processed when DCBx is configured to operate in CEE mode and vice versa. In this case, the unrecognized TLVs cause the unrecognized TLV counter to increment, but the frame is processed and is not discarded.

Legacy DCBx (CIN and CEE) supports the DCBx control state machine that is defined to maintain the sequence number and acknowledge the number sent in the DCBx control TLVs.

Behavior of Tagged Packets

The below is example for enabling PFC for priority 2 for tagged packets. Priority (Packet Dot1p) 2 will be mapped to PG6 on PRIO2PG setting. All other Priorities for which PFC is not enabled are mapped to default PG – PG7.

Classification rules on ingress (Ingress FP CAM region) matches incoming packet-dot1p and assigns an internal priority (to select queue as per Table 1 and Table 2).

The internal Priority assigned for the packet by Ingress FP is used by the memory management unit (MMU) to assign the packet to right queue by indexing the internal-priority to queue map table (TABLE 1) in hardware.

PRIO2COS setting for honoring the PFC protocol packets from the Peer switches is as per above Packet-Dot1p->queue table (Table 2).

The packets that come in with packet-dot1p 2 alone will be assigned to PG6 on ingress.

The packets that come in with packet-dot1p 2 alone will use Q1 (as per dot1p to Queue classification – Table 2) on the egress port.

- When Peer sends a PFC message for Priority 2, based on above PRIO2COS table (TABLE 2), Queue 1 is halted.
- Queue 1 starts buffering the packets with Dot1p 2. This causes PG6 buffer counter to increase on the ingress, since P-dot1p 2 is mapped to PG6.
- As the PG6 watermark threshold is reached, PFC will be generated for dot1p 2.

Configuration Example for DSCP and PFC Priorities

Consider a scenario in which the following DSCP and PFC priorities are necessary:

DSCP	0 – 5, 10 - 15	20 – 25, 30 – 35
Expected PFC Priority	1	2

To configure the aforementioned DSCP and PFC priority values, perform the following tasks:

1. Create class-maps to group the DSCP subsets

```
class-map match-any dscp-pfc-1
 match ip dscp 0-5,10-15
!
class-map match-any dscp-pfc-2
 match ip dscp 20-25,30-35
```

2. Associate above class-maps to Queues Queue assignment to be based on the below table (This internal table has been modified for S6000 platforms).

Table 11.

Internal-priority	0	1	2	3	4	5	6	7
Queue	1	2	0	3	4	5	6	7

```
policy-map-input policy-dscp-based-pfc
 service-queue 0 class-map dscp-pfc-1
 service-queue 1 class-map dscp-pfc-2
```

3. Dot1p->Queue Mapping Configuration is retained at the default value.

```
Default dot1p-queue mapping is,
Dell#show qos dot1p-queue-mapping
 Dot1p Priority : 0  1  2  3  4  5  6  7
              Queue :2  0  1  3  4  5  6  7
```

4. Interface Configurations on server connected ports.

   a. Enable DCB globally.

   ```
   Dell(conf)#dcb enable
   ```

   b. Apply Policy-map configuration. Configure the policy-map ‘policy-dscp-based-pfc’ on the server connected untagged ports.

   ```
   Dell(conf)#int te 0/0
   Dell(conf-if-te-0/0)#service-policy input policy-dscp-based-pfc
   ```

   c. Apply PFC Priority configuration. Configure priorities on which PFC is enabled.

   ```
   Dell(conf-if-te-0/0)#pfc priority 1,2
   ```

DCBx Example

The following figure shows how to use DCBx.

The external 40GbE ports on the base module (ports 33 and 37) of two switches are used for uplinks configured as DCBx auto-upstream ports. The is connected to third-party, top-of-rack (ToR) switches through 40GbE uplinks. The ToR switches are part of a Fibre Channel storage network.

The internal ports (ports 1-32) connected to the 10GbE backplane are configured as auto-downstream ports.

On the , PFC and ETS use DCBx to exchange link-level configuration with DCBx peer devices.

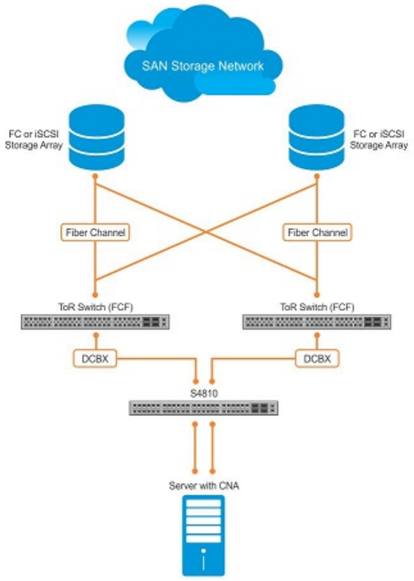

SAN Storage Network
FC or iSCSI Storage Array
FC or iSCSI Storage Array
Fiber Channel
Fiber Channel
ToR Switch (FCF)
ToR Switch (FCF)
DCBX
DCBX
S4810
Server with CNA

Figure 29. DCBx Sample Topology

DCBx Prerequisites and Restrictions

The following prerequisites and restrictions apply when you configure DCBx operation on a port:

- For DCBx, on a port interface, enable LLDP in both Send (TX) and Receive (RX) mode (the `protocol lldp mode` command; refer to the example in CONFIGURATION versus INTERFACE Configurations in the Link Layer Discovery Protocol (LLDP) chapter). If multiple DCBx peer ports are detected on a local DCBx interface, LLDP is shut down.
- The CIN version of DCBx supports only PFC, ETS, and FCOE; it does not support iSCSI, backward congestion management (BCN), logical link down (LLDF), and network interface virtualization (NIV).

Configuring DCBx

To configure DCBx, follow these steps.
For DCBx, to advertise DCBx TLVs to peers, enable LLDP. For more information, refer to Link Layer Discovery Protocol (LLDP).
Configure DCBx operation at the interface level on a switch or globally on the switch. To configure the system for DCBx operation in a data center network, you must:

1. Configure ToR- and FCF-facing interfaces as auto-upstream ports.
2. Configure server-facing interfaces as auto-downstream ports.
3. Configure a port to operate in a configuration-source role.
4. Configure ports to operate in a manual role.

1. Enter INTERFACE Configuration mode.

CONFIGURATION mode

```
interface type slot/port
```

2. Enter LLDP Configuration mode to enable DCBx operation.

INTERFACE mode

```
[no] protocol lldp
```

3. Configure the DCBx version used on the interface, where: `auto` configures the port to operate using the DCBx version received from a peer.

   PROTOCOL LLDP mode

   ```
   [no] DCBx version {auto | cee | cin | ieee-v2.5}
   ```

   - `cee`: configures the port to use CEE (Intel 1.01).
   - `cin`: configures the port to use Cisco-Intel-Nuova (DCBx 1.0).
   - `ieee-v2.5`: configures the port to use IEEE 802.1Qaz (Draft 2.5).

   The default is **Auto**.

4. Configure the DCBx port role the interface uses to exchange DCB information.

   PROTOCOL LLDP mode

   ```
   [no] DCBx port-role {config-source | auto-downstream | auto-upstream |
   manual}
   ```

   - `auto-upstream`: configures the port to receive a peer configuration. The configuration source is elected from auto-upstream ports.
   - `auto-downstream`: configures the port to accept the internally propagated DCB configuration from a configuration source.
   - `config-source`: configures the port to serve as the configuration source on the switch.
   - `manual`: configures the port to operate only on administer-configured DCB parameters. The port does not accept a DCB configuration received from a peer or a local configuration source.

   The default is **Manual**.

5. **On manual ports only**: Configure the PFC and ETS TLVs advertised to DCBx peers.

   PROTOCOL LLDP mode

   ```
   [no] advertise DCBx-tlv {ets-conf | ets-reco | pfc} [ets-conf | ets-reco |
   pfc] [ets-conf | ets-reco | pfc]
   ```

   - `ets-conf`: enables the advertisement of ETS Configuration TLVs.
   - `ets-reco`: enables the advertisement of ETS Recommend TLVs.
   - `pfc enables`: the advertisement of PFC TLVs.

   The default is All PFC and ETS TLVs are advertised.

   NOTE: You can configure the transmission of more than one TLV type at a time; for example, advertise `DCBx-tlv ets-conf ets-reco`. You can enable ETS recommend TLVs (`ets-reco`) only if you enable ETS configuration TLVs (`ets-conf`).

   To disable TLV transmission, use the `no` form of the command; for example, `no advertise DCBx-tlv pfc ets-reco`.

6. **On manual ports only**: Configure the Application Priority TLVs advertised on the interface to DCBx peers.

   PROTOCOL LLDP mode

   ```
   [no] advertise DCBx-appln-tlv {fcoe | iscsi}
   ```

- `fcoe`: enables the advertisement of FCoE in Application Priority TLVs.
- `iscsi`: enables the advertisement of iSCSI in Application Priority TLVs.

The default is Application Priority TLVs are enabled to advertise FCoE and iSCSI.

NOTE: To disable TLV transmission, use the `no` form of the command; for example, `no advertise DCBx-appln-tlv iscsi`.

For information about how to use iSCSI, refer to iSCSI Optimization.

To verify the DCBx configuration on a port, use the `show interface DCBx detail` command.

Configuring DCBx Globally on the Switch

To globally configure the DCBx operation on a switch, follow these steps.

1. Enter Global Configuration mode.
   EXEC PRIVILEGE mode

   ```
   configure
   ```
2. Enter LLDP Configuration mode to enable DCBx operation.
   CONFIGURATION mode

   ```
   [no] protocol lldp
   ```
3. Configure the DCBx version used on all interfaces not already configured to exchange DCB information.
   PROTOCL LLDP mode

   ```
   [no] DCBx version {auto | cee | cin | ieee-v2.5}
   ```

   - `auto`: configures all ports to operate using the DCBx version received from a peer.
   - `cee`: configures a port to use CEE (Intel 1.01). cin configures a port to use Cisco-Intel-Nuova (DCBx 1.0).
   - `ieee-v2.5`: configures a port to use IEEE 802.1Qaz (Draft 2.5).

   The default is **Auto**.

   NOTE: To configure the DCBx port role the interfaces use to exchange DCB information, use the `DCBx port-role` command in INTERFACE Configuration mode (Step 3).
4. Configure the PFC and ETS TLVs that advertise on unconfigured interfaces with a manual port-role.
   PROTOCOL LLDP mode

   ```
   [no] advertise DCBx-tlv {ets-conf | ets-reco | pfc} [ets-conf | ets-reco |
   pfc] [ets-conf | ets-reco | pfc]
   ```

   - `ets-conf`: enables transmission of ETS Configuration TLVs.
   - `ets-reco`: enables transmission of ETS Recommend TLVs.
   - `pfc`: enables transmission of PFC TLVs.

NOTE: You can configure the transmission of more than one TLV type at a time. You can only enable ETS recommend TLVs (`ets-reco`) if you enable ETS configuration TLVs (`ets-conf`). To disable TLV transmission, use the `no` form of the command; for example, `no advertise DCBx-tlv pfc ets-reco`.

The default is All TLV types are enabled.

5. Configure the Application Priority TLVs that advertise on unconfigured interfaces with a manual port-role.
   PROTOCOL LLDP mode

   ```
   [no] advertise DCBx-appln-tlv {fcoe | iscsi}
   ```

   - `fcoe`: enables the advertisement of FCoE in Application Priority TLVs.
   - `iscsi`: enables the advertisement of iSCSI in Application Priority TLVs.

   The default is Application Priority TLVs are enabled and advertise FCoE and iSCSI.

   NOTE: To disable TLV transmission, use the `no` form of the command; for example, `no advertise DCBx-appln-tlv iscsi`.

   For information about how to use iSCSI, refer to iSCSI Optimization.
6. Configure the FCoE priority advertised for the FCoE protocol in Application Priority TLVs.
   PROTOCOL LLDP mode

   ```
   [no] fcoe priority-bits priority-bitmap
   ```

   The priority-bitmap range is from 1 to FF.

   The default is **0x8**.
7. Configure the iSCSI priority advertised for the iSCSI protocol in Application Priority TLVs.
   PROTOCOL LLDP mode

   ```
   [no] iscsi priority-bits priority-bitmap
   ```

   The priority-bitmap range is from 1 to FF.

   The default is **0x10**.

DCBx Error Messages

The following syslog messages appear when an error in DCBx operation occurs.

```
LLDP_MULTIPLE_PEER_DETECTED: DCBx is operationally disabled after detecting
more than one DCBx
peer on the port interface.
```

```
LLDP_PEER_AGE_OUT: DCBx is disabled as a result of LLDP timing out on a DCBx
peer interface.
```

```
DSM_DCBx_PEER_VERSION_CONFLICT: A local port expected to receive the IEEE, CIN,
or CEE version
```

```
in a DCBx TLV from a remote peer but received a different, conflicting DCBx
version.
```

```
DSM_DCBx_PFC_PARAMETERS_MATCH and DSM_DCBx_PFC_PARAMETERS_MISMATCH: A local
DCBx port received
a compatible (match) or incompatible (mismatch) PFC configuration from a peer.
```

```
DSM_DCBx_ETS_PARAMETERS_MATCH and DSM_DCBx_ETS_PARAMETERS_MISMATCH: A local
DCBx port received
a compatible (match) or incompatible (mismatch) ETS configuration from a peer.
```

```
LLDP_UNRECOGNISED_DCBx_TLV_RECEIVED: A local DCBx port received an unrecognized
DCBx TLV from
a peer.
```

Debugging DCBx on an Interface

To enable DCBx debug traces for all or a specific control paths, use the following command.

- Enable DCBx debugging.
  EXEC PRIVILEGE mode

  ```
  debug DCBx {all | auto-detect-timer | config-exchng | fail | mgmt | resource
  | sem | tlv}
  ```

  - `all`: enables all DCBx debugging operations.
  - `auto-detect-timer`: enables traces for DCBx auto-detect timers.
  - `config-exchng`: enables traces for DCBx configuration exchanges.
  - `fail`: enables traces for DCBx failures.
  - `mgmt`: enables traces for DCBx management frames.
  - `resource`: enables traces for DCBx system resource frames.
  - `sem`: enables traces for the DCBx state machine.
  - `tlv`: enables traces for DCBx TLVs.

Verifying the DCB Configuration

To display DCB configurations, use the following `show` commands.

Table 12. Displaying DCB Configurations

Command	Output
`show dot1p-queue mapping`	Displays the current 802.1p priority-queue mapping.
`show dcb [stack-unit` *`unit-number`*`]`	Displays the data center bridging status, number of PFC-enabled ports, and number of PFC-enabled queues. On the master switch in a stack, you can specify a stack-unit number. The range is from 0 to 5.
`show qos dcb-input [pfc-profile]`	Displays the PFC configuration in a DCB input policy.
`show qos dcb-output [ets-profile]`	Displays the ETS configuration in a DCB output policy.

Command	Output
`show qos priority-groups`	Displays the ETS priority groups configured on the switch, including the 802.1p priority classes and ID of each group.
`show interface` *`port-type slot/port`* `pfc {summary | detail}`	Displays the PFC configuration applied to ingress traffic on an interface, including priorities and link delay. To clear PFC TLV counters, use the `clear pfc counters interface` *`port-type slot/port`* command.
`show interface` *`port-type slot/port`* `pfc statistics`	Displays counters for the PFC frames received and transmitted (by dot1p priority class) on an interface.
`show interface` *`port-type slot/port`* `ets {summary | detail}`	Displays the ETS configuration applied to egress traffic on an interface, including priority groups with priorities and bandwidth allocation. To clear ETS TLV counters, enter the `clear ets counters interface` *`port-type slot/port`* command.
`show interface` *`port-type slot/port`* `DCBx detail`	Plays the DCBx configuration on an interface.
`show stack-unit {`*`0-11`*` | all} stack ports all pfc details`	Displays the PFC configuration applied to ingress traffic on stack-links, including priorities and link delay.
`show stack-unit {`*`0-11`*` | all} stack ports all ets details`	Displays the ETS configuration applied to ingress traffic on stack-links, including priorities and link delay.

Examples of the `show` Commands

The following example shows the `show dot1p-queue mapping` command.

```
Dell(conf)# show dot1p-queue-mapping
Dot1p Priority: 0 1 2 3 4 5 6 7
Queue         : 0 0 0 1 2 3 3 3
```

The following example shows the `show dcb` command.

```
Dell# show dcb
stack-unit 0 port-set 0
     DCB Status : Enabled
 PFC Port Count : 56 (current), 56 (configured)
PFC Queue Count : 2 (current),  2 (configured)
```

The following example shows the `show qos dcb-input` command.

```
Dell(conf)# show qos dcb-input
dcb-input pfc-profile
  pfc link-delay 32
  pfc priority 0-1
dcb-input pfc-profile1
  no pfc mode on
  pfc priority 6-7
```

The following example shows the `show qos dcb-output` command.

```
Dell# show qos dcb-output
dcb-output ets
  priority-group san qos-policy san
  priority-group ipc qos-policy ipc
  priority-group lan qos-policy lan
```

The following example shows the `show qos priority-groups` command.

```
Dell#show qos priority-groups
priority-group ipc
  priority-list 4
  set-pgid 2
```

The following example shows the `show interfaces pfc summary` command.

```
Dell# show interfaces tengigabitethernet 0/49 pfc summary
Interface TenGigabitEthernet 0/49
    Admin mode is on
    Admin is enabled
    Remote is enabled, Priority list is 4
    Remote Willing Status is enabled
    Local is enabled
    Oper status is Recommended
    PFC DCBx Oper status is Up
    State Machine Type is Feature
    TLV Tx Status is enabled
    PFC Link Delay 45556 pause quantams
    Application Priority TLV Parameters :
    --------------------------------------
    FCOE TLV Tx Status is disabled
    ISCSI TLV Tx Status is disabled
    Local FCOE PriorityMap is 0x8
    Local ISCSI PriorityMap is 0x10
    Remote FCOE PriorityMap is 0x8
    Remote ISCSI PriorityMap is 0x8

Dell# show interfaces tengigabitethernet 0/49 pfc detail
Interface TenGigabitEthernet 0/49
    Admin mode is on
    Admin is enabled
    Remote is enabled
    Remote Willing Status is enabled
    Local is enabled
    Oper status is recommended
    PFC DCBx Oper status is Up
    State Machine Type is Feature
    TLV Tx Status is enabled
    PFC Link Delay 45556 pause quanta
    Application Priority TLV Parameters :
    --------------------------------------
    FCOE TLV Tx Status is disabled
    ISCSI TLV Tx Status is disabled
    Local FCOE PriorityMap is 0x8
    Local ISCSI PriorityMap is 0x10
    Remote FCOE PriorityMap is 0x8
    Remote ISCSI PriorityMap is 0x8

0 Input TLV pkts, 1 Output TLV pkts, 0 Error pkts, 0 Pause Tx pkts, 0 Pause Rx
pkts
```

The following table describes the `show interface pfc summary` command fields.

Table 13. **`show interface pfc summary`** Command Description

Fields	Description
Interface	Interface type with stack-unit and port number.
Admin mode is on; Admin is enabled	PFC Admin mode is on or off with a list of the configured PFC priorities . When PFC admin mode is on, PFC advertisements are enabled to be sent and received from peers; received PFC configuration takes effect. The admin operational status for a DCBx exchange of PFC configuration is enabled or disabled.
Remote is enabled; Priority list Remote Willing Status is enabled	Operational status (enabled or disabled) of peer device for DCBx exchange of PFC configuration with a list of the configured PFC priorities. Willing status of peer device for DCBx exchange (Willing bit received in PFC TLV): enabled or disabled.
Local is enabled	DCBx operational status (enabled or disabled) with a list of the configured PFC priorities
Operational status (local port)	DCBx operational status (enabled or disabled) with a list of the configured PFC priorities. Port state for current operational PFC configuration: • Init: Local PFC configuration parameters were exchanged with peer. • Recommend: Remote PFC configuration parameters were received from peer. • Internally propagated: PFC configuration parameters were received from configuration source.
PFC DCBx Oper status	Operational status for exchange of PFC configuration on local port: match (up) or mismatch (down).
State Machine Type	Type of state machine used for DCBx exchanges of PFC parameters: • Feature: for legacy DCBx versions • Symmetric: for an IEEE version
TLV Tx Status	Status of PFC TLV advertisements: enabled or disabled.
PFC Link Delay	Link delay (in quanta) used to pause specified priority traffic.
Application Priority TLV: FCOE TLV Tx Status	Status of FCoE advertisements in application priority TLVs from local DCBx port: enabled or disabled.

Fields	Description
Application Priority TLV: ISCSI TLV Tx Status	Status of ISCSI advertisements in application priority TLVs from local DCBx port: enabled or disabled.
Application Priority TLV: Local FCOE Priority Map	Priority bitmap used by local DCBx port in FCoE advertisements in application priority TLVs.
Application Priority TLV: Local ISCSI Priority Map	Priority bitmap used by local DCBx port in ISCSI advertisements in application priority TLVs.
Application Priority TLV: Remote FCOE Priority Map	Status of FCoE advertisements in application priority TLVs from remote peer port: enabled or disabled.
Application Priority TLV: Remote ISCSI Priority Map	Status of iSCSI advertisements in application priority TLVs from remote peer port: enabled or disabled.
PFC TLV Statistics: Input TLV pkts	Number of PFC TLVs received.
PFC TLV Statistics: Output TLV pkts	Number of PFC TLVs transmitted.
PFC TLV Statistics: Error pkts	Number of PFC error packets received.
PFC TLV Statistics: Pause Tx pkts	Number of PFC pause frames transmitted.
PFC TLV Statistics: Pause Rx pkts	Number of PFC pause frames received

The following example shows the `show interface pfc statistics` command.

```
Dell#show interfaces te 0/0 pfc statistics
Interface TenGigabitEthernet 0/0
Priority Received PFC Frames Transmitted PFC Frames
-------- ------------------- ----------------------
0        0                   0
1        0                   0
2        0                   0
3        0                   0
4        0                   0
5        0                   0
6        0                   0
7        0                   0
```

The following example shows the `show interface ets summary` command.

```
Dell(conf-qos-policy-out-ets)#do sho int te 0/3 ets su

Interface TenGigabitEthernet 0/3
Max Supported TC Groups is 4
Number of Traffic Classes is 8
Admin mode is on

Admin Parameters :
------------------
Admin is enabled

TC-grp   Priority#   Bandwidth   TSA
------------------------------------------------
0                    -           -
```

```
1        0,1,2       100%       ETS
2        3           0 %        SP
3        4,5,6,7     0 %        SP
4                    -          -
5                    -          -
6                    -          -
7                    -          -

Remote Parameters :
-------------------
Remote is disabled

Local Parameters :
------------------
Local is enabled

TC-grp    Priority#     Bandwidth    TSA
------------------------------------------------
0                       -            -
1         0,1,2         100%         ETS
2         3             0 %          SP
3         4,5,6,7       0 %          SP
4                       -            -
5                       -            -
6                       -            -
7                       -            -

Oper status is init
ETS DCBx Oper status is Down
State Machine Type is Asymmetric
Conf TLV Tx Status is enabled
Reco TLV Tx Status is enabled

0 Input Conf TLV Pkts, 1955 Output Conf TLV Pkts, 0 Error Conf TLV Pkts
0 Input Reco TLV Pkts, 1955 Output Reco TLV Pkts, 0 Error Reco TLV Pkts
Dell(conf)# show interfaces tengigabitethernet 0/0 ets detail
Interface TenGigabitEthernet 0/0
Max Supported TC Groups is 4
Number of Traffic Classes is 8
Admin mode is on
Admin Parameters :
------------------
Admin is enabled
TC-grp Priority# Bandwidth TSA
0 0,1,2,3,4,5,6,7 100% ETS
1 0% ETS
2 0% ETS
3 0% ETS
4 0% ETS
5 0% ETS
6 0% ETS
7 0% ETS
Priority# Bandwidth TSA
0 13% ETS
1 13% ETS
2 13% ETS
3 13% ETS
4 12% ETS
5 12% ETS
6 12% ETS
7 12% ETS
Remote Parameters:
------------------
Remote is disabled
```

```
Local Parameters :
------------------
Local is enabled
TC-grp Priority# Bandwidth TSA
0 0,1,2,3,4,5,6,7 100% ETS
1 0% ETS
2 0% ETS
3 0% ETS
4 0% ETS
5 0% ETS
6 0% ETS
7 0% ETS
Priority# Bandwidth TSA
0 13% ETS
1 13% ETS
2 13% ETS
3 13% ETS
4 12% ETS
5 12% ETS
6 12% ETS
7 12% ETS
Oper status is init
Conf TLV Tx Status is disabled
Traffic Class TLV Tx Status is disabled
0 Input Conf TLV Pkts, 0 Output Conf TLV Pkts, 0 Error Conf TLV Pkts
0T LIVnput Traffic Class TLV Pkts, 0 Output Traffic Class TLV Pkts, 0 Error
Traffic Class
Pkts
```

The following example shows the `show interface ets detail` command.

```
Dell(conf)# show interfaces tengigabitethernet 0/0 ets detail
Interface TenGigabitEthernet 0/0
Max Supported TC Groups is 4
Number of Traffic Classes is 8
Admin mode is on
Admin Parameters :
------------------
Admin is enabled
TC-grp    Priority#         Bandwidth    TSA
0         0,1,2,3,4,5,6,7   100%         ETS
1                           0%           ETS
2                           0%           ETS
3                           0%           ETS
4                           0%           ETS
5                           0%           ETS
6                           0%           ETS
7                           0%           ETS

Priority# Bandwidth TSA
0                           13%          ETS
1                           13%          ETS
2                           13%          ETS
3                           13%          ETS
4                           12%          ETS
5                           12%          ETS
6                           12%          ETS
7                           12%          ETS
Remote Parameters:
-------------------
Remote is disabled

Local Parameters :
```

```
------------------
Local is enabled
TC-grp    Priority#         Bandwidth    TSA
0         0,1,2,3,4,5,6,7   100%         ETS
1                           0%           ETS
2                           0%           ETS
3                           0%           ETS
4                           0%           ETS
5                           0%           ETS
6                           0%           ETS
7                           0%           ETS

Priority#                   Bandwidth    TSA
0                           13%          ETS
1                           13%          ETS
2                           13%          ETS
3                           13%          ETS
4                           12%          ETS
5                           12%          ETS
6                           12%          ETS
7                           12%          ETS
Oper status is init
Conf TLV Tx Status is disabled
Traffic Class TLV Tx Status is disabled
0 Input Conf TLV Pkts, 0 Output Conf TLV Pkts, 0 Error Conf TLV Pkts
0 Input Traffic Class TLV Pkts, 0 Output Traffic Class TLV Pkts, 0 Error
Traffic Class TLV
Pkts
```

The following table describes the `show interface ets detail` command fields.

Table 14. `show interface ets detail` Command Description

Field	Description
Interface	Interface type with stack-unit and port number.
Max Supported TC Group	Maximum number of priority groups supported.
Number of Traffic Classes	Number of 802.1p priorities currently configured.
Admin mode	ETS mode: on or off. When on, the scheduling and bandwidth allocation configured in an ETS output policy or received in a DCBx TLV from a peer can take effect on an interface.
Admin Parameters	ETS configuration on local port, including priority groups, assigned dot1p priorities, and bandwidth allocation.
Remote Parameters	ETS configuration on remote peer port, including Admin mode (enabled if a valid TLV was received or disabled), priority groups, assigned dot1p priorities, and bandwidth allocation. If the ETS Admin mode is enabled on the remote port for DCBx exchange, the Willing bit received in ETS TLVs from the remote peer is included.

Field	Description
Local Parameters	ETS configuration on local port, including Admin mode (enabled when a valid TLV is received from a peer), priority groups, assigned dot1p priorities, and bandwidth allocation.
Operational status (local port)	Port state for current operational ETS configuration: • Init: Local ETS configuration parameters were exchanged with peer. • Recommend: Remote ETS configuration parameters were received from peer. • Internally propagated: ETS configuration parameters were received from configuration source.
ETS DCBx Oper status	Operational status of ETS configuration on local port: match or mismatch.
State Machine Type	Type of state machine used for DCBx exchanges of ETS parameters: • Feature: for legacy DCBx versions • Asymmetric: for an IEEE version
Conf TLV Tx Status	Status of ETS Configuration TLV advertisements: enabled or disabled.
ETS TLV Statistic: Input Conf TLV pkts	Number of ETS Configuration TLVs received.
ETS TLV Statistic: Output Conf TLV pkts	Number of ETS Configuration TLVs transmitted.
ETS TLV Statistic: Error Conf TLV pkts	Number of ETS Error Configuration TLVs received.

The following example shows the `show stack-unit all stack-ports all pfc details` command.

```
Dell(conf)# show stack-unit all stack-ports all pfc details

stack unit 0 stack-port all
    Admin mode is On
    Admin is enabled, Priority list is 4-5
    Local is enabled, Priority list is 4-5
    Link Delay 45556 pause quantum
    0 Pause Tx pkts, 0 Pause Rx pkts

stack unit 1 stack-port all
    Admin mode is On
    Admin is enabled, Priority list is 4-5
    Local is enabled, Priority list is 4-5
    Link Delay 45556 pause quantum
    0 Pause Tx pkts, 0 Pause Rx pkts
```

The following example shows the `show stack-unit all stack-ports all ets details` command.

```
Dell(conf)# show stack-unit all stack-ports all ets details
Stack unit 0 stack port all
Max Supported TC Groups is 4
```

```
Number of Traffic Classes is 1

Admin mode is on
Admin Parameters:
--------------------
Admin is enabled
TC-grp    Priority#        Bandwidth    TSA
------------------------------------------------
0         0,1,2,3,4,5,6,7  100%         ETS
1                          -            -
2                          -            -
3                          -            -
4                          -            -
5                          -            -
6                          -            -
7                          -            -
8                          -            -

Stack unit 1 stack port all
Max Supported TC Groups is 4
Number of Traffic Classes is 1
Admin mode is on
Admin Parameters:
--------------------
Admin is enabled
TC-grp    Priority#        Bandwidth    TSA
------------------------------------------------
0         0,1,2,3,4,5,6,7  100%         ETS
1                          -            -
2                          -            -
3                          -            -
4                          -            -
5                          -            -
6                          -            -
7                          -            -
8                          -            -
```

The following example shows the `show interface DCBx detail` command (IEEE).

```
Dell(conf-if-te-0/17-lldp)#do sho int te 2/12 dc d

E-ETS Configuration TLV enabled             e-ETS Configuration TLV disabled
R-ETS Recommendation TLV enabled            r-ETS Recommendation TLV disabled
P-PFC Configuration TLV enabled             p-PFC Configuration TLV disabled
F-Application priority for FCOE enabled     f-Application Priority for FCOE
disabled
I-Application priority for iSCSI enabled    i-Application Priority for iSCSI
disabled
--------------------------------------------------------------------------------
----------

Interface TenGigabitEthernet 2/12
  Remote Mac Address 00:01:e8:8a:df:a0
    Port Role is Manual
  DCBx Operational Status is Enabled
  Is Configuration Source? FALSE
  Local DCBx Compatibility mode is IEEEv2.5
  Local DCBx Configured mode is IEEEv2.5
  Peer Operating version is IEEEv2.5
  Local DCBx TLVs Transmitted: ERPFi
  1 Input PFC TLV pkts, 2 Output PFC TLV pkts, 0 Error PFC pkts
  0 PFC Pause Tx pkts, 0 Pause Rx pkts
  1 Input ETS Conf TLV Pkts, 1 Output ETS Conf TLV Pkts, 0 Error ETS Conf TLV
```

```
Pkts
  1 Input ETS Reco TLV pkts, 1 Output ETS Reco TLV pkts, 0 Error ETS Reco TLV
Pkts
```

The following example shows the `show interface DCBx detail` command (legacy CEE).

```
Dell(conf-if-te-0/17-lldp)#do sho int te 1/14 dc d

E-ETS Configuration TLV enabled             e-ETS Configuration TLV disabled
R-ETS Recommendation TLV enabled            r-ETS Recommendation TLV disabled
P-PFC Configuration TLV enabled             p-PFC Configuration TLV disabled
F-Application priority for FCOE enabled     f-Application Priority for FCOE
disabled
I-Application priority for iSCSI enabled    i-Application Priority for iSCSI
disabled
-----------------------------------------------------------------------

Interface TenGigabitEthernet 1/14
  Remote Mac Address 00:01:e8:8a:df:a0
    Port Role is Auto-Upstream
  DCBx Operational Status is Enabled
  Is Configuration Source? FALSE
  Local DCBx Compatibility mode is CEE
  Local DCBx Configured mode is CEE
  Peer Operating version is CEE
  Local DCBx TLVs Transmitted: ErPFi

Local DCBx Status
-----------------
  DCBx Operational Version is 0
  DCBx Max Version Supported is 0
  Sequence Number: 1
  Acknowledgment Number: 1
  Protocol State: In-Sync

Peer DCBx Status:
----------------
  DCBx Operational Version is 0
  DCBx Max Version Supported is 0
  Sequence Number: 1
  Acknowledgment Number: 1
  Total DCBx Frames transmitted 994
  Total DCBx Frames received 646
  Total DCBx Frame errors 0
  Total DCBx Frames unrecognized 0
```

The following table describes the `show interface DCBx detail` command fields.

Table 15. `show interface DCBx detail` Command Description

Field	Description
Interface	Interface type with chassis slot and port number.
Port-Role	Configured DCBx port role: auto-upstream, auto-downstream, config-source, or manual.
DCBx Operational Status	Operational status (enabled or disabled) used to elect a configuration source and internally propagate a DCB configuration. The DCBx operational status is the combination of PFC and ETS operational status.

Field	Description
Configuration Source	Specifies whether the port serves as the DCBx configuration source on the switch: true (yes) or false (no).
Local DCBx Compatibility mode	DCBx version accepted in a DCB configuration as compatible. In auto-upstream mode, a port can only received a DCBx version supported on the remote peer.
Local DCBx Configured mode	DCBx version configured on the port: CEE, CIN, IEEE v2.5, or Auto (port auto-configures to use the DCBx version received from a peer).
Peer Operating version	DCBx version that the peer uses to exchange DCB parameters.
Local DCBx TLVs Transmitted	Transmission status (enabled or disabled) of advertised DCB TLVs (see TLV code at the top of the show command output).
Local DCBx Status: DCBx Operational Version	DCBx version advertised in Control TLVs.
Local DCBx Status: DCBx Max Version Supported	Highest DCBx version supported in Control TLVs.
Local DCBx Status: Sequence Number	Sequence number transmitted in Control TLVs.
Local DCBx Status: Acknowledgment Number	Acknowledgement number transmitted in Control TLVs.
Local DCBx Status: Protocol State	Current operational state of DCBx protocol: ACK or IN-SYNC.
Peer DCBx Status: DCBx Operational Version	DCBx version advertised in Control TLVs received from peer device.
Peer DCBx Status: DCBx Max Version Supported	Highest DCBx version supported in Control TLVs received from peer device.
Peer DCBx Status: Sequence Number	Sequence number transmitted in Control TLVs received from peer device.
Peer DCBx Status: Acknowledgment Number	Acknowledgement number transmitted in Control TLVs received from peer device.
Total DCBx Frames transmitted	Number of DCBx frames sent from local port.
Total DCBx Frames received	Number of DCBx frames received from remote peer port.
Total DCBx Frame errors	Number of DCBx frames with errors received.
Total DCBx Frames unrecognized	Number of unrecognizable DCBx frames received.

Generation of PFC for a Priority for Untagged Packets

In order to generate PFC for a particular priority for untagged packets, and configuring PFC for that priority, you should find the queue number associated with priority from TABLE 1 and Associate an input policy map to forward the matched DSCP packet to that queue. PFC frames gets generated with PFC priority associated with the queue when the queue gets congested.

Operations on Untagged Packets

The below is example for enabling PFC for priority 2 for tagged packets. Priority (Packet Dot1p) 2 will be mapped to PG6 on PRIO2PG setting. All other Priorities for which PFC is not enabled are mapped to default PG – PG7.

Classification rules on ingress (Ingress FP CAM region) matches incoming packet-dot1p and assigns an internal priority (to select queue as per Table 1 and Table 2).

The internal Priority assigned for the packet by Ingress FP is used by the memory management unit (MMU) to assign the packet to right queue by indexing the internal-priority to queue map table (TABLE 1) in hardware.

PRIO2COS setting for honoring the PFC protocol packets from the Peer switches is as per above Packet-Dot1p->queue table (Table 2).

The packets that come in with packet-dot1p 2 alone will be assigned to PG6 on ingress.

The packets that come in with packet-dot1p 2 alone will use Q1 (as per dot1p to Queue classification – Table 2) on the egress port.

- When Peer sends a PFC message for Priority 2, based on above PRIO2COS table (TABLE 2), Queue 1 is halted.
- Queue 1 starts buffering the packets with Dot1p 2. This causes PG6 buffer counter to increase on the ingress, since P-dot1p 2 is mapped to PG6.
- As the PG6 watermark threshold is reached, PFC will be generated for dot1p 2.

Performing PFC Using DSCP Bits Instead of 802.1p Bits

Priority based Flow Control (PFC) is currently supported on Dell Networking OS for tagged packets based on the packet Dot1p. In certain data center deployments, VLAN configuration is avoided on the servers and all packets from the servers are untagged. These packets will carry IP header and can be differentiated based on the DSCP fields they carry on the server facing switch ports. Requirement is to classify these untagged packets from the server based on their DSCP and provide PFC treatment.

Dell Networking OS Releases 9.3(0.0) and earlier provide CLI support to specify the priorities for which PFC is enabled on each port. This feature is applicable only for the tagged packets based on the incoming packet Dot1p and Dot1p based queue classification. This document will discuss the configurations required to support PFC for untagged packets based on incoming packet DSCP.

For the tagged packets, Queue is selected based on the incoming Packet Dot1p. When PFC frames for a specific priority is received from the peer switch, the queue corresponding to that Dot1p is halted from scheduling on that port, thus honoring the PFC from the peer. If a queue is congested due to packets with a specific Dot1p and PFC is enabled for that Dot1p, switch will transit out PFC frames for that Dot1p. The packet Dot1p to Queue mapping for classification on the ingress must be same as the mapping of Dot1p to the Queue to be halted on the egress used for PFC honoring. Dell Networking OS ensures that these mappings are identical. This section discusses the Dell Networking OS configurations needed for above PFC generation and honoring mechanism to work for the untagged packets.

PRIORITY to PG mapping (PRIO2PG) is on the ingress for each port. By default, all priorities are mapped to PG7. A priority for which PFC has to be generated is assigned to a PG other than PG7 (say PG6) and buffer watermark is set on PG6 so as to generate PFC.

In ingress, the buffers are accounted at per PG basis and would indicate the number of the packets that has ingress this port PG but still queued up in egress pipeline. However, there is no direct mapping between the PG and Queue.

Packet is assigned an internal priority on the ingress pipeline based on the queue to which it is destined. This Internal-priority to Queue mapping has been modified and enhanced as follows for S6000:

Table 16. Priority to Queue Mapping

Internal-priority	0	1	2	3	4	5	6	7
Queue	2	0	1	3	4	5	6	7

Default dot1p to queue configuration is as follows:

Table 17. Dot1p to Queue Mapping

Packet-Dot1p	0	1	2	3	4	5	6	7
Queue	2	0	1	3	4	5	6	7

PFC and ETS Configuration Examples

This section contains examples of how to configure and apply DCB input and output policies on an interface.

Using PFC and ETS to Manage Data Center Traffic

The following shows examples of using PFC and ETS to manage your data center traffic.

In the following example:

- Incoming SAN traffic is configured for priority-based flow control.
- Outbound LAN, IPC, and SAN traffic is mapped into three ETS priority groups and configured for enhanced traffic selection (bandwidth allocation and scheduling).
- One lossless queue is used.

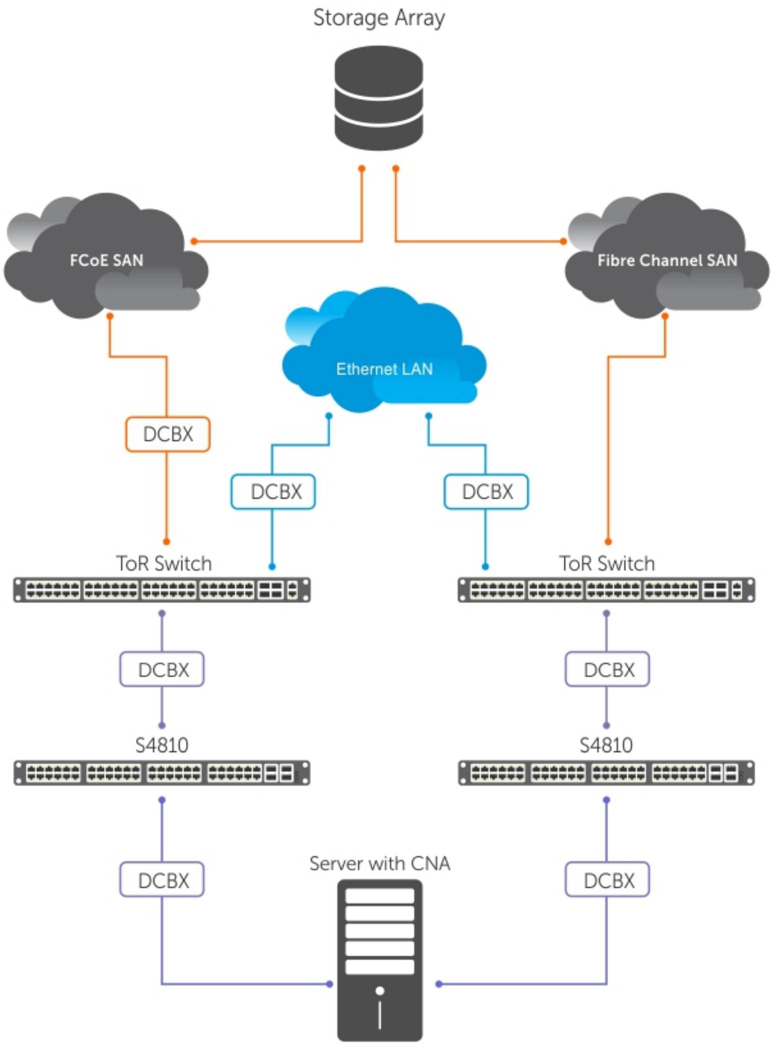

Storage Array
FCoE SAN
Fibre Channel SAN
Ethernet LAN
DCBX
DCBX
DCBX
DCBX
ToR Switch
ToR Switch
DCBX
DCBX
S4810
S4810
DCBX
Server with CNA
DCBX

Figure 30. PFC and ETS Applied to LAN, IPC, and SAN Priority Traffic

QoS Traffic Classification: The `service-class dynamic dot1p` command has been used in Global Configuration mode to map ingress dot1p frames to the queues shown in the following table. For more information, refer to QoS dot1p Traffic Classification and Queue Assignment.

The following describes the dot1p-priority class group assignment

dot1p Value in the Incoming Frame	Priority Group Assignment
0	LAN
1	LAN
2	LAN

dot1p Value in the Incoming Frame	Priority Group Assignment
3	SAN
4	IPC
5	LAN
6	LAN
7	LAN

The following describes the priority group-bandwidth assignment.

Priority Group	Bandwidth Assignment
IPC	5%
SAN	50%
LAN	45%

PFC and ETS Configuration Command Examples

The following examples show PFC and ETS configuration commands to manage your data center traffic.

Example of Configuring QoS Priority-Queue Assignment to Honor Dot1p Priorities

```
Dell(conf)# service-class dynamic dot1p

Or

Dell(conf)# interface tengigabitethernet 0/1
Dell(conf-if-te-0/1)# service-class dynamic dot1p
```

Example of Configuring a DCB Input Policy to Apply PFC to Lossless SAN Priority Traffic

```
Dell(conf)# dcb-input ipc_san_lan
Dell(conf-qos-policy-in)# pfc mode on
Dell(conf-qos-policy-in)# pfc priority 3
```

Example of Configuring an ETS Priority Group

```
Dell(conf)# priority-group san
Dell(conf-pg)# priority-list 3
Dell(conf-pg)# set-pgid 1
Dell(conf-pg)# exit
Dell(conf)# priority-group ipc
Dell(conf-pg)# priority-list 4
Dell(conf-pg)# set-pgid 2
Dell(conf-pg)# exit
Dell(conf)# priority-group lan
Dell(conf-pg)# priority-list 0-2,5-7
Dell(conf-pg)# set-pgid 3
Dell(conf-pg)# exit
```

Example of Configuring an ETS Output Policy for Egress Traffic

```
Dell(conf)# qos-policy-output san ets
Dell(conf-qos-policy-out)# bandwidth-percentage 50
Dell(conf-qos-policy-out)# exit
Dell(conf)# qos-policy-output lan ets
Dell(conf-qos-policy-out)# bandwidth-percentage 45
Dell(conf-qos-policy-out)# exit
Dell(conf)# qos-policy-output ipc ets
Dell(conf-qos-policy-out)# bandwidth-percentage 5
Dell(conf-qos-policy-out)# exit
```

Example of Configuring a DCB Output Policy to Apply ETS (Bandwidth Allocation and Scheduling) to IPC, SAN, and LAN Priority Traffic

```
Dell(conf)# dcb-output ets
Dell(conf-dcb-out)# priority-group san qos-policy san
Dell(conf-dcb-out)# priority-group lan qos-policy lan
Dell(conf-dcb-out)# priority-group ipc qos-policy ipc
```

Example of Applying DCB Input and Output Policies to an Interface

```
Dell(conf)# interface tengigabitethernet 0/1
Dell(conf-if-te-0/1)# dcb-policy input pfc
Dell(conf-if-te-0/1)# dcb-policy output ets
```

Example of Configuring a QoS Output Policy to Specify Bandwidth Allocation to Different Traffic Types if DCBx Version is CIN

```
Dell(conf)# qos-policy-output lan-q0
Dell(conf-qos-policy-out)# bandwidth-percentage 20
Dell(conf-qos-policy-out)# exit
Dell(conf)#q os-policy-output lan-q3
Dell(conf-qos-policy-out)# bandwidth-percentage 70
Dell(conf-qos-policy-out)# exit
Dell(conf)#policy-map-output ets-queues
```

Example of Creating a QoS Policy Map for DCBx CIN Bandwidth Allocation

```
Dell(conf)# policy-map-output ets-queues
Dell(conf-policy-map-out)# service-queue 0 qos-policy lan-q0
Dell(conf-policy-map-out)# service-queue 3 qos-policy lan-q3
```

Example of Applying the QoS Policy Map for DCBx CIN Bandwidth Allocation to an Interface

```
Dell(conf-if-te-0/1)# service-policy output ets-queues
```

Using PFC and ETS to Manage Converged Ethernet Traffic in a Switch Stack

The following example shows how to apply the DCB PFC input policy (`ipc_san_lan`) and ETS output policy (`ets`) on all ports in a switch stack.

This example references the PFC and ETS Configuration Examples section.

Example of Applying DCB PFC Input Policy and ETS Output Policy in a Switch Stack

```
Dell(conf)# dcb-policy output stack-unit all stack-ports all ets
Dell(conf)# dcb-policy input stack-unit all stack-ports all pfc
```

Hierarchical Scheduling in ETS Output Policies

ETS supports up to three levels of hierarchical scheduling.

For example, you can apply ETS output policies with the following configurations:

Priority group 1	Assigns traffic to one priority queue with 20% of the link bandwidth and strict-priority scheduling.
Priority group 2	Assigns traffic to one priority queue with 30% of the link bandwidth.
Priority group 3	Assigns traffic to two priority queues with 50% of the link bandwidth and strict-priority scheduling.

In this example, the configured ETS bandwidth allocation and scheduler behavior is as follows:

Unused bandwidth usage:	Normally, if there is no traffic or unused bandwidth for a priority group, the bandwidth allocated to the group is distributed to the other priority groups according to the bandwidth percentage allocated to each group. However, when three priority groups with different bandwidth allocations are used on an interface: • If priority group 3 has free bandwidth, it is distributed as follows: 20% of the free bandwidth to priority group 1 and 30% of the free bandwidth to priority group 2. • If priority group 1 or 2 has free bandwidth, (20 + 30)% of the free bandwidth is distributed to priority group 3. Priority groups 1 and 2 retain whatever free bandwidth remains up to the (20+ 30)%.
Strict-priority groups:	If two priority groups have strict-priority scheduling, traffic assigned from the priority group with the higher priority-queue number is scheduled first. However, when three priority groups are used and two groups have strict-priority scheduling (such as groups 1 and 3 in the example), the strict priority group whose traffic is mapped to one queue takes precedence over the strict priority group whose traffic is mapped to two queues.

Therefore, in this example, scheduling traffic to priority group 1 (mapped to one strict-priority queue) takes precedence over scheduling traffic to priority group 3 (mapped to two strict-priority queues).

Configuring DCB Maps and its Attributes

This topic contains the following sections that describe how to configure a DCB map, apply the configured DCB map to a port, configure PFC without a DCB map, and configure lossless queues. This functionality is supported on the S6000 platform.

DCB Map: Configuration Procedure

A DCB map consists of PFC and ETS parameters. By default, PFC is not enabled on any 802.1p priority and ETS allocates equal bandwidth to each priority. To configure user-defined PFC and ETS settings, you must create a DCB map.

Step	Task	Command	Command Mode
1	Enter global configuration mode to create a DCB map or edit PFC and ETS settings.	`dcb-map` *`name`*	CONFIGURATION
2	Configure the PFC setting (on or off) and the ETS bandwidth percentage allocated to traffic in each priority group, or whether the priority group traffic should be handled with strict priority scheduling. You can enable PFC on a maximum of two priority queues on an interface. Enabling PFC for dot1p priorities makes the corresponding port queue lossless. The sum of all allocated bandwidth percentages in all groups in the DCB map must be 100%. Strict-priority traffic is serviced first. Afterwards, bandwidth allocated to other	`priority-group` *`group_num`* {**bandwidth** *percentage* | **strict-priority**} **pfc** {**on** | **off**}	DCB MAP

Step	Task	Command	Command Mode
	priority groups is made available and allocated according to the specified percentages. If a priority group does not use its allocated bandwidth, the unused bandwidth is made available to other priority groups. Example: **priority-group 0 bandwidth 60 pfc off priority-group 1 bandwidth 20 pfc on priority-group 2 bandwidth 20 pfc on priority-group 4 strict-priority pfc off** Repeat this step to configure PFC and ETS traffic handling for each priority group.		
3	Specify the dot1p priority-to-priority group mapping for each priority. Priority-group range: 0 to 7. All priorities that map to the same queue must be in the same priority group. Leave a space between each priority group number. For example: **priority-pgid 0 0 0 1 2 4 4 4** in which priority group 0 maps to dot1p priorities 0, 1, and 2; priority group 1 maps to dot1p priority 3; priority group 2 maps to dot1p priority 4; priority group 4 maps to dot1p priorities 5, 6, and 7.	`priority-pgid` *`dot1p0_group_num`* *`dot1p1_group_num`* *`dot1p2_group_num`* *`dot1p3_group_num`* *`dot1p4_group_num`* *`dot1p5_group_num`* *`dot1p6_group_num`* *`dot1p7_group_num`*	DCB MAP

Important Points to Remember

- If you remove a dot1p priority-to-priority group mapping from a DCB map (`no priority pgid` command), the PFC and ETS parameters revert to their default values on the interfaces on which the DCB map is applied. By default, PFC is not applied on specific 802.1p priorities; ETS assigns equal bandwidth to each 802.1p priority.

  As a result, PFC and lossless port queues are disabled on 802.1p priorities, and all priorities are mapped to the same priority queue and equally share the port bandwidth.
- To change the ETS bandwidth allocation configured for a priority group in a DCB map, do not modify the existing DCB map configuration. Instead, first create a new DCB map with the desired PFC and ETS settings, and apply the new map to the interfaces to override the previous DCB map settings. Then, delete the original dot1p priority-priority group mapping.

  If you delete the dot1p priority-priority group mapping (`no priority pgid` command) before you apply the new DCB map, the default PFC and ETS parameters are applied on the interfaces. This change may create a DCB mismatch with peer DCB devices and interrupt network operation.

Applying a DCB Map on a Port

When you apply a DCB map with PFC enabled on an S6000 interface, a memory buffer for PFC-enabled priority traffic is automatically allocated. The buffer size is allocated according to the number of PFC-enabled priorities in the assigned map.

To apply a DCB map to an Ethernet port, follow these steps:

Step	Task	Command	Command Mode
1	Enter interface configuration mode on an Ethernet port.	`interface` {**tengigabitEthernet** *`slot/port`* | **fortygigabitEthernet** *`slot/port`*}	CONFIGURATION
2	Apply the DCB map on the Ethernet port to configure it with the PFC and ETS settings in the map; for example: Dell# **interface tengigabitEthernet 0/0** Dell(config-if-te-0/0)# **dcb-map SAN_A_dcb_map1** Repeat Steps 1 and 2 to apply a DCB map to more than one port. You cannot apply a DCB map on an interface that has been already configured for PFC using the`pfc priority` command or which is already configured for lossless queues (`pfc no-drop queues` command).	`dcb-map` *`name`*	INTERFACE

Configuring PFC without a DCB Map

In a network topology that uses the default ETS bandwidth allocation (assigns equal bandwidth to each priority), you can also enable PFC for specific dot1p-priorities on individual interfaces without using a DCB map. This type of DCB configuration is useful on interfaces that require PFC for lossless traffic, but do not transmit converged Ethernet traffic.

Step	Task	Command	Command Mode
1	Enter interface configuration mode on an Ethernet port.	`interface {tengigabitEthernet` *`slot`***/port |** `fortygigabitEthernet` *`slot`***/port}**	CONFIGURATION
2	Enable PFC on specified priorities. Range: 0-7. Default: None. Maximum number of lossless queues supported on an Ethernet port: 2. Separate priority values with a comma. Specify a priority range with a dash, for example: pfc priority 3,5-7 1. You cannot configure PFC using the `pfc priority` command on an interface on which a DCB map has been applied or which is already configured for lossless queues (`pfc no-drop queues` command).	`pfc priority` *`priority-range`*	INTERFACE

Configuring Lossless Queues

DCB also supports the manual configuration of lossless queues on an interface after you disable PFC mode in a DCB map and apply the map on the interface. The configuration of no-drop queues provides flexibility for ports on which PFC is not needed, but lossless traffic should egress from the interface.

Lossless traffic egresses out the no-drop queues. Ingress 802.1p traffic from PFC-enabled peers is automatically mapped to the no-drop egress queues.

When configuring lossless queues on a port interface, consider the following points:

- By default, no lossless queues are configured on a port.
- A limit of two lossless queues are supported on a port. If the number of lossless queues configured exceeds the maximum supported limit per port (two), an error message is displayed. You must re-configure the value to a smaller number of queues.
- If you configure lossless queues on an interface that already has a DCB map with PFC enabled (**pfc on**), an error message is displayed.

Step	Task	Command	Command Mode
1	Enter INTERFACE Configuration mode.	`interface`{**tengigabitEthernet** *`slot/port`* `|` **fortygigabitEthernet** *`slot/port`*}	CONFIGURATION
2	Open a DCB map and enter DCB map configuration mode.	`dcb-map` *`name`*	INTERFACE
3	Disable PFC.	`no pfc mode on`	DCB MAP
4	Return to interface configuration mode.	`exit`	DCB MAP
5	Apply the DCB map, created to disable the PFC operation, on the interface	`dcb-map` {*name* | **default**}	INTERFACE
6	Configure the port queues that still function as no-drop queues for lossless traffic. For the dot1p-queue assignments. The maximum number of lossless queues globally supported on a port is 2. You cannot configure PFC no-drop queues on an interface on which a DCB map with PFC enabled has been applied, or which is already configured for PFC using the `pfc priority` command. Range: 0-3. Separate queue values with a comma; specify a priority range with a dash; for example: pfc no-drop queues 1,3 or pfc no-drop queues 2-3 Default: No lossless queues are configured.	`pfc no-drop queues`*`queue-range`*	INTERFACE

Data Center Bridging: Default Configuration

This functionality is supported on the S6000 platform.

Before you configure PFC and ETS on an S6000 switch (see Configuring DCB Maps and its Attributes), take into account the following default settings:

DCB is enabled.

The PFC memory buffer supports up to 52 (not 64) PFC-enabled ports and two lossless queues per port.

PFC and ETS are globally enabled by default.

The default dot1p priority-queue assignments are applied as follows:

```
Dell(conf)#do show qos dot1p-queue-mapping
 Dot1p Priority : 0  1  2  3  4  5  6  7
          Queue : 2  0  1  3  4  5  6  7
Dell(conf)#
```

NOTE: In Egress queue assignment (8 queues in S6000, 4 against in S5000 / S4810.

PFC is not applied on specific dot1p priorities.

ETS: Equal bandwidth is assigned to each port queue and each dot1p priority in a priority group.

To configure PFC and ETS parameters on an S6000 interface, you must specify the PFC mode, the ETS bandwidth allocation for a priority group, and the 802.1p priority-to-priority group mapping in a DCB map. No default PFC and ETS settings are applied to Ethernet interfaces.

Configuring PFC and ETS in a DCB Map

An S6000 switch supports the use of a DCB map in which you configure priority-based flow control (PFC) and enhanced transmission selection (ETS) settings. To configure PFC and ETS parameters, you must apply a DCB map on an S6000 interface. This functionality is supported on the S6000 platform.

PFC Configuration Notes

PFC provides flow control based on the 802.1p priorities in a converged Ethernet traffic that is received on an interface and is enabled by default when you enable DCB. As an enhancement to the existing Ethernet pause functionality, PFC stops traffic transmission for specified priorities (CoS values) without impacting other priority classes. Different traffic types are assigned to different priority classes.

When traffic congestion occurs, PFC sends a pause frame to a peer device with the CoS priority values of the traffic that needs to be stopped. DCBx provides the link-level exchange of PFC parameters between peer devices. PFC allows network administrators to create zero-loss links for SAN traffic that requires no-drop service, while at the same time retaining packet-drop congestion management for LAN traffic.

On an S6000 switch, PFC is enabled by default on Ethernet ports (`pfc mode on` command). You can configure PFC parameters using a DCB map or the `pfc priority` command in Interface configuration mode. For more information, see Configuring DCB Maps and its Attributes.

As soon as you apply a DCB map with PFC enabled on an interface, DCBx starts exchanging information with a peer. The IEEE802.1Qbb, CEE and CIN versions of PFC TLV are supported. DCBx also validates PFC configurations that are received in TLVs from peer devices. By applying a DCB map with PFC enabled, you enable PFC operations on ingress port traffic. To achieve complete lossless handling of traffic, configure PFC priorities on all DCB egress ports.

NOTE: DCB maps are supported only on physical Ethernet interfaces.

- To remove a DCB map, including the PFC configuration it contains, use the `no dcb map` command in Interface configuration mode.
- To disable PFC operation on an interface, use the `no pfc mode on` command in DCB-Map configuration mode.
- Traffic may be interrupted when you reconfigure PFC no-drop priorities in a DCB map or re-apply the DCB map to an interface.
- For PFC to be applied, the configured priority traffic must be supported by a PFC peer (as detected by DCBx).
- If you apply a DCB map with PFC disabled (`pfc off`), you can enable link-level flow control on the interface using the `flowcontrol rx on tx on` command. To delete the DCB map, first disable link-level flow control. PFC is then automatically enabled on the interface because an interface is PFC-enabled by default.
- To ensure no-drop handling of lossless traffic, PFC allows you to configure lossless queues on a port (see Configuring DCB Maps and its Attributes).
- When you configure a DCB map, an error message is displayed if the PFC dot1p priorities result in more than two lossless queues.
- When you apply a DCB map, an error message is displayed if link-level flow control is already enabled on an interface. You cannot enable PFC and link-level flow control at the same time on an interface.
- In a switch stack, configure all stacked ports with the same PFC configuration.
- Dell Networking OS allows you to change the default dot1p priority-queue assignments only if the change satisfies the following requirements in DCB maps already applied to S6000 interfaces:
- All 802.1p priorities mapped to the same queue must be in the same priority group.
- A maximum of two PFC-enabled, lossless queues are supported on an interface.

Otherwise, the reconfiguration of a default dot1p-queue assignment is rejected.

- To ensure complete no-drop service, apply the same PFC parameters on all PFC-enabled peers.

PFC Prerequisites and Restrictions

On an S6000 switch, PFC is globally enabled by default, but not applied on specific 802.1p priorities. To enable PFC on 802.1p priorities, create a DCB map. For more information, see Configuring DCB Maps and its Attributes.

The following prerequisites and restrictions apply when you configure PFC in a DCB map:

- You can enable PFC on a maximum of two priority queues on an interface. Enabling PFC for dot1p priorities configures the corresponding port queue as lossless.
- You cannot enable PFC and link-level flow control at the same time on an interface.

ETS Configuration Notes

ETS provides a way to optimize bandwidth allocation to outbound 802.1p classes of converged Ethernet traffic. Different traffic types have different service needs. Using ETS, you can create groups within an 802.1p priority class to configure different treatment for traffics with different bandwidth, latency, and best-effort needs.

When you configure ETS in a DCB map:

- The DCB map associates a priority group with a PFC operational mode (on or off) and an ETS scheduling and bandwidth allocation. You can apply a DCB map on multiple egress ports.
- Use the ETS configuration associated with 802.1p priority traffic in a DCB map in DCBx negotiation with ETS peers.
- Traffic in priority groups is assigned to strict-queue or weighted round-robin (WRR) scheduling in an ETS configuration and is managed using the ETS bandwidth-assignment algorithm. Dell Networking OS de-queues all frames of strict-priority traffic before servicing any other queues. A queue with strict-priority traffic can starve other queues in the same port.
- ETS-assigned bandwidth allocation and strict-priority scheduling apply only to data queues, not to control queues.
- Dell Networking OS supports hierarchical scheduling on an interface. The control traffic on Dell Networking OS is redirected to control queues as higher priority traffic with strict priority scheduling. After the control queues drain out, the remaining data traffic is scheduled to queues according to the bandwidth and scheduler configuration in the DCB map. The available bandwidth calculated by the ETS algorithm is equal to the link bandwidth after scheduling non-ETS higher-priority traffic.
- The configuration of bandwidth allocation and strict-queue scheduling is not supported at the same time for a priority group.
- **Bandwidth assignment**: By default, equal bandwidth is assigned to each dot1p priority in a priority group. To configure the bandwidth assigned to the port queues associated with dot1p priorities in a priority group, use the **bandwidth** *percentage* parameter. The sum of the bandwidth allocated to all priority groups in a DCB map must be 100% of the bandwidth on the link. You must allocate at least 1% of the total bandwidth to each priority group.
- **Scheduling of priority traffic**: dot1p priority traffic on the switch is scheduled to the current queue mapping. dot1p priorities within the same queue must have the same traffic properties and scheduling method.
- **ETS configuration error**: If an error occurs in an ETS configuration, the configuration is ignored and the scheduler and bandwidth allocation settings are reset to the ETS default value: 100% of available bandwidth is allocated to priority group 0 and the bandwidth is equally assigned to each dot1p priority.

If an error occurs when a port receives a peer’s ETS configuration, the port’s configuration resets to the ETS configuration in the previously configured DCB map. If no DCB map was previously applied, the port resets to the default ETS parameters.

ETS Prerequisites and Restrictions

On an S6000 switch, ETS is enabled by default on Ethernet ports with equal bandwidth assigned to each 802.1p priority. You can change the default ETS configuration only by using a DCB map. For more information, see Configuring DCB Maps and its Attributes.

The following prerequisites and restrictions apply when you configure ETS bandwidth allocation or strict-priority queuing in a DCB map:

- When allocating bandwidth or configuring strict-priority queuing for dot1p priorities in a priority group on a DCBx CIN interface, take into account the CIN bandwidth allocation (see Configuring Bandwidth Allocation for DCBx CIN) and dot1p-queue mapping.
- Although ETS bandwidth allocation or strict-priority queuing does not support weighted random early detection (WRED), explicit congestion notification (ECN), rate shaping, and rate limiting because these parameters are not negotiated by DCBx with peer devices, you can apply a QoS output policy with WRED and/or rate shaping on a DCBx CIN-enabled interface. In this case, the WRED or rate shaping configuration in the QoS output policy must take into account the bandwidth allocation or queue scheduler configured in the DCB map.

Priority-Group Configuration Notes

When you configure priority groups in a DCB map:

- A priority group consists of 802.1p priority values that are grouped together for similar bandwidth allocation and scheduling, and that share the same latency and loss requirements. All 802.1p priorities mapped to the same queue must be in the same priority group.
- In a DCB map, each 802.1p priority must map to a priority group.
- The maximum number of priority groups supported in a DCB map on an interface is equal to the number of data queues (4) on the port. Each priority group can support more than one data queue.
- You can enable PFC on a maximum of two priority queues on an interface.
- If you configure more than one priority group as strict priority, the higher numbered priority queue is given preference when scheduling data traffic.

Priority-Based Flow Control Using Dynamic Buffer Method

Priority-based flow control using dynamic buffer spaces is supported on the S6000 platform.

In a data center network, priority-based flow control (PFC) manages large bursts of one traffic type in multiprotocol links so that it does not affect other traffic types and no frames are lost due to congestion. When PFC detects congestion on a queue for a specified priority, it sends a pause frame for the 802.1p priority traffic to the transmitting device.

Pause and Resume of Traffic

The pause message is used by the sending device to inform the receiving device about a congested, heavily-loaded traffic state that has been identified. When the interface of a sending device transmits a pause frame, the recipient acknowledges this frame by temporarily halting the transmission of data packets. The sending device requests the recipient to restart the transmission of data traffic when the congestion eases and reduces. The time period that is specified in the pause frame defines the duration for which the flow of data packets is halted. When the time period elapses, the transmission restarts.

When a device sends a pause frame to another device, the time for which the sending of packets from the other device must be stopped is contained in the pause frame. The device that sent the pause frame empties the buffer to be less than the threshold value and restarts the acceptance of data packets.

Dynamic ingress buffering enables the sending of pause frames at different thresholds based on the number of ports that experience congestion at a time. This behavior impacts the total buffer size used by a particular lossless priority on an interface. The pause and resume thresholds can also be configured dynamically. You can configure a buffer size, pause threshold, ingress shared threshold weight, and resume threshold to control and manage the total amount of buffers that are to be used in your network environment.

All the PFC-related settings such as the DCB input and output policies or DCB maps are saved in the DCB application and the Differentiated Services Manager (DSM) application. All of these configurations can be modified only for interfaces that are enabled for DCB. The DCB buffer configurations are also saved in the DCB and DSM databases.

Buffer Sizes for Lossless or PFC Packets

You can configure up to a maximum of 4 lossless (PFC) queues. By configuring 4 lossless queues, you can configure 4 different priorities and assign a particular priority to each application that your network is used to process. For example, you can assign a higher priority for time-sensitive applications and a lower priority for other services, such as file transfers. You can configure the amount of buffer space to be allocated for each priority and the pause or resume thresholds for the buffer. This method of configuration enables you to effectively manage and administer the behavior of lossless queues.

Although the system contains 9 MB of space for shared buffers, a minimum guaranteed buffer is provided to all the internal and external ports in the system for both unicast and multicast traffic. This minimum guaranteed buffer reduces the total available shared buffer to 7,787 KB. This shared buffer can be used for lossy and lossless traffic.

The default behavior causes up to a maximum of 6.6 MB to be used for PFC-related traffic. The remaining approximate space of 1 MB can be used by lossy traffic. You can allocate all the remaining 1 MB to lossless PFC queues. If you allocate in such a way, the performance of lossy traffic is reduced and degraded. Although you can allocate a maximum buffer size, it is used only if a PFC priority is configured and applied on the interface.

The number of lossless queues supported on the system is dependent on the availability of total buffers for PFC. The default configuration in the system guarantees a minimum of 52 KB per queue if all the 128 queues are congested. However, modifying the buffer allocation per queue impacts this default behavior.

By default the total available buffer for PFC is 6.6 MB and when you configure dynamic ingress buffering, a minimum of least 52 KB per queue is used when all ports are congested. By default, the system enables a maximum of 1 lossless queue on the S6000 platform.

This default behavior is impacted if you modify the total buffer available for PFC or assign static buffer configurations to the individual PFC queues.

Interworking of DCB Map With DCB Buffer Threshold Settings

DCB map functionality is supported on the S6000 platform.

The `dcb-input` and `dcb-output` configuration commands are deprecated. You must use the `dcp-map` command to create a DCB map to configure priority flow control (PFC) and enhanced transmission selection (ETS) on Ethernet ports that support converged Ethernet traffic.

Configure the `dcb-buffer-threshold` command and its related parameters only on ports with either auto configuration or dcb-map configuration. This command is not supported on existing front-panel interfaces or stack ports that are configured with the `dcb-input` or `dcb-output` commands. Similarly, if the dcb-buffer-threshold configuration is present on a stack port or any interface, the dcb-input or dcb-ouput policies cannot be applied on those interfaces.

Example: When the dcb-buffer-threshold policy is applied on interfaces or stack ports with the dcb-input or dcb-output policies, the following error message is displayed:

```
%Error: dcb-buffer-threshold not supported on interfaces with deprecated
commands
```

Example: When the dcb-input or dcb-output policy is configured on interfaces or stack ports with the dcb-buffer threshold policy, the following error message is displayed:

```
%Error: Deprecated command is not supported on interfaces with dcb-buffer-
threshold configured
```

You must not modify the service-class dot1p mappings when any buffer-threshold-policy is configured on the system.

```
S4810-1(conf)#service-class dot1p-mapping dot1p0 3

% Error: PFC buffer-threshold policies conflict with dot1p mappings. Please
remove all dcb-buffer-threshold policies to change mappings.
```

The `show dcb` command has been enhanced to display the following additional buffer-related information:

```
 S4810-YU-MR-Dell (conf)#do show dcb
 dcb Status :  Enabled
 PFC Queue Count :  2 --Indicate the PFC queue configured.
 Total buffer (lossy + lossless)(in KB): 7787--Total buffer space for lossy and
lossless queues
 PFC total buffer (in KB): 6526 --Indicates the total buffer (configured or
default)
 PFC shared buffer (in KB): 832--Indicates the shared buffer (Configured or
default)
 PFC available buffer ( in KB): 5694--Indicates remaining available buffers for
PFC that are free to be allocated
```

Configuring the Dynamic Buffer Method

Priority-based flow control using dynamic buffer spaces is supported on the S6000 platform.

To configure the dynamic buffer capability, perform the following steps:

1. Enable the DCB application. By default, DCB is enabled and link-level flow control is disabled on all interfaces.

   CONFIGURATION mode

   ```
   S6000-109-Dell(conf)#dcb enable
   ```

2. Configure the shared PFC buffer size and the total buffer size. A maximum of 4 lossless queues are supported.

   CONFIGURATION mode

```
S6000-109-Dell(conf)#dcb pfc-shared-buffer-size 4000
```

```
S6000-109-Dell(conf)#dcb pfc-total-buffer-size 5000
```

3. Configure the number of PFC queues.

   CONFIGURATION mode

   ```
   Dell(conf)#dcb enable pfc-queues 4
   ```

   The number of ports supported based on lossless queues configured will depend on the buffer. The default number of PFC queues in the system is 2 for S4810 and 1 for S6000 platforms.

   For each priority, you can specify the shared buffer threshold limit, the ingress buffer size, buffer limit for pausing the acceptance of packets, and the buffer offset limit for resuming the acceptance of received packets.

4. Configure the profile name for the DCB buffer threshold

   CONFIGURATION mode

   ```
   S6000-109-Dell(conf)#dcb-buffer-threshold test
   ```

5. DCB-BUFFER-THRESHOLD mode

   ```
   S4810-YU-MR-Dell(conf-dcb-buffer-thr)# priority 0 buffer-size 52 pause-
   threshold 16 resume-offset 10 shared-threshold-weight 7
   ```

6. Assign the DCB policy to the DCB buffer threshold profile on stack ports.

   CONFIGURATION mode

   ```
   S4810-YU-MR-Dell(conf)# dcb-policy buffer-threshold stack-unit all stack-
   ports all test
   ```

7. Assign the DCB policy to the DCB buffer threshold profile on interfaces. This setting takes precedence over the default buffer-threshold setting.

   INTERFACE mode (conf-if-te)

   ```
   S4810-YU-MR-Dell(conf-if-te-0/0)#dcb-policy buffer-threshold test
   ```

8. Create a QoS policy buffer and enter the QoS Policy Buffer Configuration mode to configure the no-drop queues, ingress buffer size, buffer limit for pausing, and buffer offset limit for resuming.

   CONFIGURATION mode

   ```
   S4810-YU-MR-Dell(conf)# qos-policy-buffer test
   ```

   ```
   S4810-YU-MR-Dell (conf-qos-policy-buffer)#queue 0 pause no-drop buffer-size
   128000 pause-threshold 103360 resume-threshold 83520
   ```

   ```
   S4810-YU-MR-Dell (conf-qos-policy-buffer)# queue 4 pause no-drop buffer-size
   128000 pause-threshold 103360 resume-threshold 83520
   ```

Applying a DCB Map in a Switch Stack

Apply the same DCB map with PFC and ETS configuration to all stacked ports in a switch stack. You cannot apply different DCB maps to different stacked switches. This functionality is supported on the S6000 platform.

To remove all PFC and ETS settings applied to stacked ports from the DCB map and reset PFC and ETS to their default settings, use the `no dcb-map stack-unit all` command.

Task	Command	Command Mode
Apply the specified DCB map on all ports of the switch stack.	`dcb-map stack-unit all stack-ports all` *`dcb-map-name`*	CONFIGURATION

12

Dynamic Host Configuration Protocol (DHCP)

Dynamic host configuration protocol (DHCP) is available on the S6000 platform.

DHCP is an application layer protocol that dynamically assigns IP addresses and other configuration parameters to network end-stations (hosts) based on configuration policies determined by network administrators.

DHCP relieves network administrators of manually configuring hosts, which can be a tedious and error-prone process when hosts often join, leave, and change locations on the network and it reclaims IP addresses that are no longer in use to prevent address exhaustion.

DHCP is based on a client-server model. A host discovers the DHCP server and requests an IP address, and the server either leases or permanently assigns one. There are three types of devices that are involved in DHCP negotiation:

DHCP Server	This is a network device offering configuration parameters to the client.
DHCP Client	This is a network device requesting configuration parameters from the server.
Relay Agent	This is an intermediary network device that passes DHCP messages between the client and server when the server is not on the same subnet as the host.

DHCP Packet Format and Options

DHCP uses the user datagram protocol (UDP) as its transport protocol.

The server listens on port 67 and transmits to port 68; the client listens on port 68 and transmits to port 67. The configuration parameters are carried as options in the DHCP packet in Type, Length, Value (TLV) format; many options are specified in RFC 2132. To limit the number of parameters that servers must provide, hosts specify the parameters that they require, and the server sends only those parameters. Some common options are shown in the following illustration.

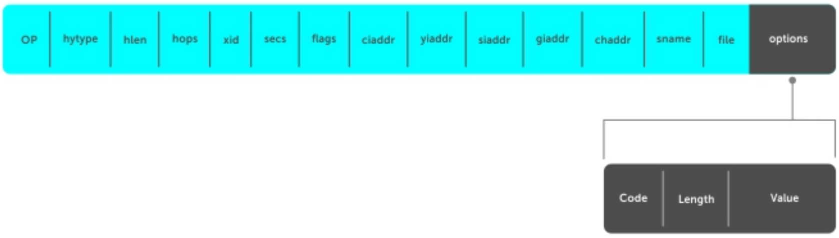


Figure 31. DHCP packet Format

The following table lists common DHCP options.

Option	Number and Description
Subnet Mask	Option 1 Specifies the client's subnet mask.
Router	Option 3 Specifies the router IP addresses that may serve as the client's default gateway.
Domain Name Server	Option 6 Specifies the domain name servers (DNSs) that are available to the client.
Domain Name	Option 15 Specifies the domain name that clients should use when resolving hostnames via DNS.
IP Address Lease Time	Option 51 Specifies the amount of time that the client is allowed to use an assigned IP address.
DHCP Message Type	Option 53 • 1: DHCPDISCOVER • 2: DHCPOFFER • 3: DHCPREQUEST • 4: DHCPDECLINE • 5: DHCPACK • 6: DHCPNACK • 7: DHCPRELEASE • 8: DHCPINFORM
Parameter Request List	Option 55 Clients use this option to tell the server which parameters it requires. It is a series of octets where each octet is DHCP option code.
Renewal Time	Option 58 Specifies the amount of time after the IP address is granted that the client attempts to renew its lease with the *original* server.
Rebinding Time	Option 59 Specifies the amount of time after the IP address is granted that the client attempts to renew its lease with *any* server, if the original server does not respond.
Vendor Class Identifer	Option 60

Option	Number and Description
	Identifiers a user-defined string used by the Relay Agent to forward DHCP client packets to a specific server.
L2 DHCP Snooping	Option 82 Specifies IP addresses for DHCP messages received from the client that are to be monitored to build a DHCP snooping database.
User Port Stacking	Option 230 Set the stacking option variable to provide DHCP server stack-port detail when the DHCP offer is set.
End	Option 255 Signals the last option in the DHCP packet.

Assign an IP Address using DHCP

The following section describes DHCP and the client in a network.

When a client joins a network:

1. The client initially broadcasts a **DHCPDISCOVER** message on the subnet to discover available DHCP servers. This message includes the parameters that the client requires and might include suggested values for those parameters.
2. Servers unicast or broadcast a **DHCPOFFER** message in response to the DHCPDISCOVER that offers to the client values for the requested parameters. Multiple servers might respond to a single DHCPDISCOVER; the client might wait a period of time and then act on the most preferred offer.
3. The client broadcasts a **DHCPREQUEST** message in response to the offer, requesting the offered values.
4. After receiving a DHCPREQUEST, the server binds the clients' unique identifier (the hardware address plus IP address) to the accepted configuration parameters and stores the data in a database called a binding table. The server then broadcasts a **DHCPACK** message, which signals to the client that it may begin using the assigned parameters.
5. When the client leaves the network, or the lease time expires, returns its IP address to the server in a **DHCPRELEASE** message.

There are additional messages that are used in case the DHCP negotiation deviates from the process previously described and shown in the illustration below.

DHCPDECLINE	A client sends this message to the server in response to a DHCPACK if the configuration parameters are unacceptable; for example, if the offered address is already in use. In this case, the client starts the configuration process over by sending a DHCPDISCOVER.
DHCPINFORM	A client uses this message to request configuration parameters when it assigned an IP address manually rather than with DHCP. The server responds by unicast.
DHCPNAK	A server sends this message to the client if it is not able to fulfill a DHCPREQUEST; for example, if the requested address is already in use. In this case, the client starts the configuration process over by sending a DHCPDISCOVER.

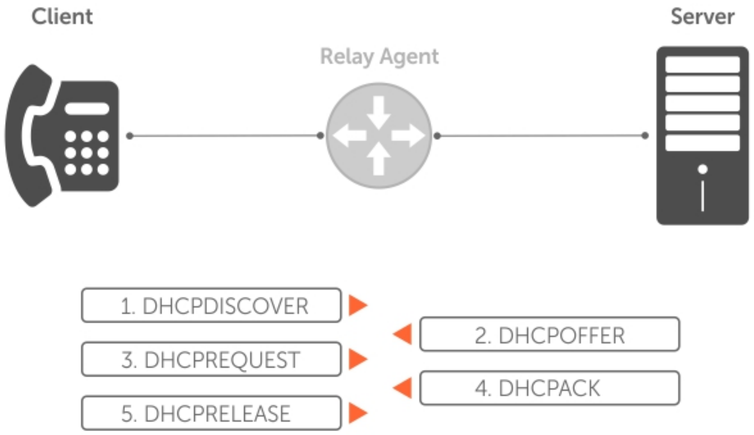


Figure 32. Client and Server Messaging

Implementation Information

The following describes DHCP implementation.

- Dell Networking implements DHCP based on RFC 2131 and RFC 3046.
- IP source address validation is a sub-feature of DHCP Snooping; the Dell Networking OS uses access control lists (ACLs) internally to implement this feature and as such, you cannot apply ACLs to an interface which has IP source address validation. If you configure IP source address validation on a member port of a virtual local area network (VLAN) and then attempt to apply an access list to the VLAN, Dell Networking OS displays the first line in the following message. If you first apply an ACL to a VLAN and then attempt enable IP source address validation on one of its member ports, Dell Networking OS displays the second line in the following message.

  ```
  % Error: Vlan member has access-list configured.
  % Error: Vlan has an access-list configured.
  ```

  NOTE: If you enable DHCP Snooping globally and you have any configured L2 ports, any IP ACL, MAC ACL, or DHCP source address validation ACL does not block DHCP packets.

- Dell Networking OS provides 40K entries that can be divided between leased addresses and excluded addresses. By extension, the maximum number of pools you can configure depends on the subnet mask that you give to each pool. For example, if all pools were configured for a /24 mask, the total would be 40000/253 (approximately 158). If the subnet is increased, more pools can be configured. The maximum subnet that can be configured for a single pool is /17. Dell Networking OS displays an error message for configurations that exceed the allocated memory.
- The S6000 platform supports 4K DHCP Snooping entries.
- All platforms support Dynamic ARP Inspection on 16 VLANs per system. For more information, refer to Dynamic ARP Inspection.

  NOTE: If the DHCP server is on the top of rack (ToR) and the VLTi (ICL) is down due to a failed link, when a VLT node is rebooted in BMP (Bare Metal Provisioning) mode, it is not able to reach the DHCP server, resulting in BMP failure.

Configure the System to be a DHCP Server

Configuring the system to be a DHCP server is supported only on the S6000 platform.

A DHCP server is a network device that has been programmed to provide network configuration parameters to clients upon request. Servers typically serve many clients, making host management much more organized and efficient.

The following table lists the key responsibilities of DHCP servers.

Table 18. DHCP Server Responsibilities

DHCP Server Responsibility	Description
Address Storage and Management	DHCP servers are the owners of the addresses used by DHCP clients.The server stores the addresses and manages their use, keeping track of which addresses have been allocated and which are still available.
Configuration Parameter Storage and Management	DHCP servers also store and maintain other parameters that are sent to clients when requested. These parameters specify in detail how a client is to operate.
Lease Management	DHCP servers use leases to allocate addresses to clients for a limited time. The DHCP server maintains information about each of the leases, including lease length.
Responding To Client Requests	DHCP servers respond to different types of requests from clients, primarily, granting, renewing, and terminating leases.
Providing Administration Services	DHCP servers include functionality that allows an administrator to implement policies that govern how DHCP performs its other tasks.

Configuring the Server for Automatic Address Allocation

Automatic address allocation is an address assignment method by which the DHCP server leases an IP address to a client from a pool of available addresses.

An address pool is a range of IP addresses that the DHCP server may assign. The subnet number indexes the address pools.

To create an address pool, follow these steps.

1. Access the DHCP server CLI context.
   CONFIGURATION mode

   ```
   ip dhcp server
   ```
2. Create an address pool and give it a name.
   DHCP mode

   ```
   pool name
   ```

3. Specify the range of IP addresses from which the DHCP server may assign addresses.

   DHCP <POOL> mode

   ```
   network network/prefix-length
   ```

   - `network`: the subnet address.
   - `prefix-length`: specifies the number of bits used for the network portion of the address you specify.

   The prefix-length range is from 17 to 31.
4. Display the current pool configuration.

   DHCP <POOL> mode

   ```
   show config
   ```

After an IP address is leased to a client, only that client may release the address. Dell Networking OS performs a IP + MAC source address validation to ensure that no client can release another clients address. This validation is a default behavior and is separate from IP+MAC source address validation.

Configuration Tasks

To configure DHCP, an administrator must first set up a DHCP server and provide it with configuration parameters and policy information including IP address ranges, lease length specifications, and configuration data that DHCP hosts need.

Configuring the Dell system to be a DHCP server is a three-step process:

1. Configuring the Server for Automatic Address Allocation
2. Specifying a Default Gateway

Related Configuration Tasks

- Configure a Method of Hostname Resolution
- Creating Manual Binding Entries
- Debugging the DHCP Server
- Using DHCP Clear Commands

Excluding Addresses from the Address Pool

The DHCP server assumes that all IP addresses in a DHCP address pool are available for assigning to DHCP clients.
You must specify the IP address that the DHCP server should not assign to clients.

To exclude an address, follow this step.

- Exclude an address range from DHCP assignment. The exclusion applies to all configured pools.

  DHCP mode

  ```
  excluded-address
  ```

Specifying an Address Lease Time

To specify an address lease time, use the following command.

- Specify an address lease time for the addresses in a pool.

  DHCP <POOL>

```
lease {days [hours] [minutes] | infinite}
```

The default is **24 hours**.

Specifying a Default Gateway

The IP address of the default router should be on the same subnet as the client.
To specify a default gateway, follow this step.

- Specify default gateway(s) for the clients on the subnet, in order of preference.

  DHCP <POOL>

  ```
  default-router address
  ```

Configure a Method of Hostname Resolution

Dell systems are capable of providing DHCP clients with parameters for two methods of hostname resolution—using DNS or NetBIOS WINS.

Using DNS for Address Resolution

A domain is a group of networks. DHCP clients query DNS IP servers when they need to correlate host names to IP addresses.

1. Create a domain.

   DHCP <POOL>

   ```
   domain-name name
   ```
2. Specify in order of preference the DNS servers that are available to a DHCP client.

   DHCP <POOL>

   ```
   dns-server address
   ```

Using NetBIOS WINS for Address Resolution

Windows internet naming service (WINS) is a name resolution service that Microsoft DHCP clients use to correlate host names to IP addresses within a group of networks. Microsoft DHCP clients can be one of four types of NetBIOS nodes: broadcast, peer-to-peer, mixed, or hybrid.

1. Specify the NetBIOS WINS name servers, in order of preference, that are available to Microsoft Dynamic Host Configuration Protocol (DHCP) clients.

   DHCP <POOL> mode

   ```
   netbios-name-server address
   ```
2. Specify the NetBIOS node type for a Microsoft DHCP client. Dell Networking recommends specifying clients as hybrid.

   DHCP <POOL> mode

   ```
   netbios-node-type type
   ```

Creating Manual Binding Entries

An address binding is a mapping between the IP address and the media access control (MAC) address of a client.
The DHCP server assigns the client an available IP address automatically, and then creates an entry in the binding table. However, the administrator can manually create an entry for a client; manual bindings are useful when you want to guarantee that a particular network device receives a particular IP address. Manual bindings can be considered single-host address pools. There is no limit on the number of manual bindings, but you can only configure one manual binding per host.

NOTE: Dell Networking OS does not prevent you from using a network IP as a host IP; be sure to not use a network IP as a host IP.

1. Create an address pool.

   DHCP mode

   ```
   pool name
   ```
2. Specify the client IP address.

   DHCP <POOL>

   ```
   host address
   ```
3. Specify the client hardware address.

   DHCP <POOL>

   ```
   hardware-address hardware-address type
   ```
   - `hardware-address`: the client MAC address.
   - `type`: the protocol of the hardware platform.

   The default protocol is **Ethernet**.

Debugging the DHCP Server

To debug the DHCP server, use the following command.

- Display debug information for DHCP server.

  EXEC Privilege mode

  ```
  debug ip dhcp server [events | packets]
  ```

Using DHCP Clear Commands

To clear DHCP binding entries, address conflicts, and server counters, use the following commands.

- Clear DHCP binding entries for the entire binding table.

  EXEC Privilege mode.

  ```
  clear ip dhcp binding
  ```
- Clear a DHCP binding entry for an individual IP address.

  EXEC Privilege mode.

  ```
  clear ip dhcp binding ip address
  ```

Configure the System to be a DHCP Client

A DHCP client is a network device that requests an IP address and configuration parameters from a DHCP server.

Implement the DHCP client functionality as follows:

- The switch can obtain a dynamically assigned IP address from a DHCP server. A start-up configuration is not received. Use bare metal provisioning (BMP) to receive configuration parameters (Dell Networking OS version and a configuration file). BMP is enabled as a factory-default setting on a switch.

  A switch cannot operate with BMP and as a DHCP client simultaneously. To disable BMP in EXEC mode, use the `stop bmp` command. After BMP stops, the switch acts as a DHCP client.
- Acquire a dynamic IP address from a DHCP client is for a limited period or until the client releases the address.
- A DHCP server manages and assigns IP addresses to clients from an address pool stored on the server. For more information, refer to Configuring the Server for Automatic Address Allocation.
- Dynamically assigned IP addresses are supported only on Ethernet interfaces: 10Gigabit, 40 Gigabit, and 100/1000/10000 Ethernet Interfaces. The DHCP client is supported on VLAN and port-channel interfaces.
- The public out-of-band management interface and default VLAN 1 are configured by default as a DHCP client to acquire a dynamic IP address from a DHCP server.
- By default, the switch is configured to operate in Jumpstart mode as a DHCP client that sends DHCP requests to a DHCP server to retrieve configuration information (IP address, boot-image filename, and configuration file). All ports and management interfaces are brought up in Layer 3 mode and pre-configured with `no shutdown` and `no ip address`. For this reason, you cannot enter configuration commands to set up the switch. To interrupt a Jumpstart process, prevent a loop from occurring, and apply the FTOS image and startup configuration stored in the local flash, enter the stop jump-start command from the console. To reconfigure the switch so that it boots up in normal mode using the FTOS image and startup configuration file in local flash, enter the reload-type normal-reload command and save it to the startup configuration:

  ```
  FTOS# reload-type normal-reload
  FTOS# write memory
  FTOS# reload
  ```
- To re-enable Jumpstart mode for the next reload, enter the `reload-type jump-start` command.

Configuring the DHCP Client System

This section describes how to configure and view an interface as a DHCP client to receive an IP address.

Dell Networking OS Behavior: The `ip address dhcp` command enables DHCP server-assigned dynamic addresses on an interface. The setting persists after a switch reboot. To stop DHCP transactions and save the dynamically acquired IP address, use the `shutdown` command on the interface. To display the dynamic IP address and show DHCP as the mode of IP address assignment, use the `show interface` *`type slot/port`* command. To unconfigure the IP address, use the `no shutdown` command when the lease timer for the dynamic IP address is expired. The interface acquires a new dynamic IP address from the DHCP server.

To configure a secondary (backup) IP address on an interface, use the `ip address` command at the INTERFACE configuration level.

Use the `no ip address dhcp` command to:

- Release the IP address dynamically acquired from a DHCP server from the interface.
- Disable the DHCP client on the interface so it cannot acquire a dynamic IP address from a DHCP server.
- Stop DHCP packet transactions on the interface.

When you enter the `release dhcp` command, the IP address dynamically acquired from a DHCP server is released from an interface. The ability to acquire a new DHCP server-assigned address remains in the running configuration for the interface. To acquire a new IP address, use the `renew DHCP` command in EXEC Privilege mode or the `ip address dhcp` command in INTERFACE Configuration mode.

To manually configure a static IP address on an interface, use the `ip address` command. A prompt displays to release an existing dynamically acquired IP address. If you confirm, the ability to receive a DHCP server-assigned IP address is removed.

To enable acquiring a dynamic IP address from a DHCP server on an interface configured with a static IP address, use the `ip address dhcp` command. A prompt displays to confirm the IP address reconfiguration. If you confirm, the statically configured IP address is released. An error message displays if you enter the `release dhcp` or `renew dhcp` commands.

To renew the lease time of the dynamically acquired IP, use the `renew dhcp` command on an interface already configured with a dynamic IP address.

NOTE: To verify the currently configured dynamic IP address on an interface, use the `show ip dhcp lease` command. The `show running-configuration` command output only displays **ip address dhcp**. The currently assigned dynamic IP address does not display.

To configure and view an interface as a DHCP client to receive an IP address, use the following commands.

1. Enter INTERFACE Configuration mode on an Ethernet interface.

   CONFIGURATION mode

   ```
   interface type slot/port
   ```
2. Acquire the IP address for an Ethernet interface from a DHCP network server.

   INTERFACE mode

   ```
   ip address dhcp
   ```

   Dynamically assigned IP addresses can be released without removing the DHCP client operation on the interface on a switch configured as a DHCP client.
3. Manually acquire a new IP address from the DHCP server by releasing a dynamically acquired IP address while retaining the DHCP client configuration on the interface.

   EXEC Privilege mode

   ```
   release dhcp interface type slot/port
   ```
4. Acquire a new IP address with renewed lease time from a DHCP server.

   EXEC Privilege mode

   ```
   renew dhcp interface type slot/port
   ```

To display DHCP client information, use the following `show` commands in EXEC Privilege mode.

- To display statistics about DHCP client interfaces, use the `show ip dhcp client statistics interface` *`type slot/port`* command.
- To clear DHCP client statistics on a specified or on all interfaces, use the `clear ip dhcp client statistics {all | interface` *`type slot/port`*`}` command.
- To display dynamic IP address lease information currently assigned to a DHCP client interface, use the `show ip dhcp lease [interface` *`type slot/port`*`]` command.
- To display log messages for all DHCP packets sent and received on DHCP client interfaces, use the `debug ip dhcp client packets [interface` *`type slot/port`*`]` command.
- To display log message on DHCP client interfaces for IP address acquisition, IP address release, IP address and lease time renewal, and release an IP address, use the `[no] debug ip dhcp client events [interface` *`type slot/port`*`]` command.

DHCP Client on a Management Interface

These conditions apply when you enable a management interface to operate as a DHCP client.

- The management default route is added with the gateway as the router IP address received in the DHCP ACK packet. It is required to send and receive traffic to and from other subnets on the external network. The route is added irrespective when the DHCP client and server are in the same or different subnets. The management default route is deleted if the management IP address is released like other DHCP client management routes.
- *ip route for 0.0.0.0* takes precedence if it is present or added later.
- Management routes added by a DHCP client display with Route Source as **DHCP** in the `show ip management route` and `show ip management-route dynamic` command output.
- Management routes added by DHCP are automatically reinstalled if you configure a static IP route with the `ip route` command that replaces a management route added by the DHCP client. If you remove the statically configured IP route using the `no ip route` command, the management route is reinstalled. Manually delete management routes added by the DHCP client.
- To reinstall management routes added by the DHCP client that is removed or replaced by the same statically configured management routes, release the DHCP IP address and renew it on the management interface.
- Management routes added by the DHCP client have higher precedence over the same statically configured management route. Static routes are not removed from the running configuration if a dynamically acquired management route added by the DHCP client overwrites a static management route.
- Management routes added by the DHCP client are not added to the running configuration.

NOTE: Management routes added by the DHCP client include the specific routes to reach a DHCP server in a different subnet and the management route.

DHCP Client Operation with Other Features

The DHCP client operates with other Dell Networking OS features, as the following describes.

Stacking

The DHCP client daemon runs only on the master unit and handles all DHCP packet transactions. It periodically synchronizes the lease file with the standby unit.

When a stack failover occurs, the new master requires the same DHCP server-assigned IP address on DHCP client interfaces. The new master reinitiates a DHCP packet transaction by sending a DHCP discovery packet on nonbound interfaces.

Virtual Link Trunking (VLT)

A DHCP client is not supported on VLT interfaces.

VLAN and Port Channels

DHCP client configuration and behavior are the same on Virtual LAN (VLAN) and port-channel (LAG) interfaces as on a physical interface.

DHCP Snooping

A DHCP client can run on a switch simultaneously with the DHCP snooping feature as follows:

- If you enable DHCP snooping globally on a switch and you enable a DHCP client on an interface, the trust port, source MAC address, and snooping table validations are not performed on the interface by DHCP snooping for packets destined to the DHCP client daemon.

  The following criteria determine packets destined for the DHCP client:

  - DHCP is enabled on the interface.
  - The user data protocol (UDP) destination port in the packet is 68.
  - The `chaddr` (change address) in the DHCP header of the packet is the same as the interface's MAC address.
- An entry in the DHCP snooping table is not added for a DHCP client interface.

DHCP Server

A switch can operate as a DHCP client and a DHCP server. DHCP client interfaces cannot acquire a dynamic IP address from the DHCP server running on the switch. Acquire a dynamic IP address from another DHCP server.

Virtual Router Redundancy Protocol (VRRP)

Do not enable the DHCP client on an interface and set the priority to 255 or assign the same DHCP interface IP address to a VRRP virtual group. Doing so guarantees that this router becomes the VRRP group owner.

To use the router as the VRRP owner, if you enable a DHCP client on an interface that is added to a VRRP group, assign a priority less than 255 but higher than any other priority assigned in the group.

Configure the System for User Port Stacking (Option 230)

Set the stacking-option variable to provide stack-port detail on the DHCP server when you set the DHCP offer. A stack can be formed when the units are connected.

Option 230 is the option for user port stacking. Use it to create up to eight stack groups. Define the configuration parameters on the DHCP server for each chassis based on the chassis MAC address. Configure the following parameters:

- unit number
- priority
- stack group ID

The received stacking configuration is always applied on the master stack unit.

```
option #230 "unit-number:3#priority:2#stack-group:14"
```

Configure Secure DHCP

The following feature is available on the platform, except where noted.

DHCP as defined by RFC 2131 provides no authentication or security mechanisms. Secure DHCP is a suite of features that protects networks that use dynamic address allocation from spoofing and attacks.

- Option 82
- DHCP Snooping
- Dynamic ARP Inspection
- Source Address Validation

Option 82

RFC 3046 (the relay agent information option, or Option 82) is used for class-based IP address assignment.
The code for the relay agent information option is 82, and is comprised of two sub-options, circuit ID and remote ID.

Circuit ID	This is the interface on which the client-originated message is received.
Remote ID	This identifies the host from which the message is received. The value of this sub-option is the MAC address of the relay agent that adds Option 82.

The DHCP relay agent inserts Option 82 before forwarding DHCP packets to the server. The server can use this information to:

- track the number of address requests per relay agent. Restricting the number of addresses available per relay agent can harden a server against address exhaustion attacks.
- associate client MAC addresses with a relay agent to prevent offering an IP address to a client spoofing the same MAC address on a different relay agent.
- assign IP addresses according to the relay agent. This prevents generating DHCP offers in response to requests from an unauthorized relay agent.

The server echoes the option back to the relay agent in its response, and the relay agent can use the information in the option to forward a reply out the interface on which the request was received rather than flooding it on the entire VLAN.

The relay agent strips Option 82 from DHCP responses before forwarding them to the client.

To insert Option 82 into DHCP packets, follow this step.

- Insert Option 82 into DHCP packets.

  CONFIGURATION mode

  ```
  ip dhcp relay information-option [trust-downstream]
  ```

  For routers between the relay agent and the DHCP server, enter the `trust-downstream` option.
- Manually reset the remote ID for Option 82.

  CONFIGURATION mode

```
ip dhcp relay information-option remote-id
```

DHCP Snooping

DHCP snooping protects networks from spoofing. In the context of DHCP snooping, ports are either trusted or not trusted.

By default, all ports are not trusted. Trusted ports are ports through which attackers cannot connect. Manually configure ports connected to legitimate servers and relay agents as trusted.

When you enable DHCP snooping, the relay agent builds a binding table — using DHCPACK messages — containing the client MAC address, IP addresses, IP address lease time, port, VLAN ID, and binding type. Every time the relay agent receives a DHCPACK on a trusted port, it adds an entry to the table.

The relay agent checks all subsequent DHCP client-originated IP traffic (DHCPRELEASE, DHCPNACK, and DHCPDECLINE) against the binding table to ensure that the MAC-IP address pair is legitimate and that the packet arrived on the correct port. Packets that do not pass this check are forwarded to the server for validation. This checkpoint prevents an attacker from spoofing a client and declining or releasing the real client's address. Server-originated packets (DHCPOFFER, DHCPACK, and DHCPNACK) that arrive on a not trusted port are also dropped. This checkpoint prevents an attacker from acting as an imposter as a DHCP server to facilitate a man-in-the-middle attack.

Binding table entries are deleted when a lease expires, or the relay agent encounters a DHCPRELEASE, DHCPNACK, or DHCPDECLINE.

Dell Networking OS Behavior: Introduced in Dell Networking OS version 7.8.1.0, DHCP snooping was available for Layer 3 only and dependent on DHCP relay agent (`ip helper-address`). Dell Networking OS version 8.2.1.0 extends DHCP snooping to Layer 2 and you do not have to enable relay agent to snoop on Layer 2 interfaces.

Dell Networking OS Behavior: Binding table entries are deleted when a lease expires or when the relay agent encounters a DHCPRELEASE. Line cards maintain a list of snooped VLANs. When the binding table is exhausted, DHCP packets are dropped on snooped VLANs, while these packets are forwarded across non-snooped VLANs. Because DHCP packets are dropped, no new IP address assignments are made. However, DHCPRELEASE and DHCPDECLINE packets are allowed so that the DHCP snooping table can decrease in size. After the table usage falls below the maximum limit of 4000 entries, new IP address assignments are allowed.

NOTE: DHCP server packets are dropped on all not trusted interfaces of a system configured for DHCP snooping. To prevent these packets from being dropped, configure `ip dhcp snooping trust` on the server-connected port.

Enabling DHCP Snooping

To enable DHCP snooping, use the following commands.

1. Enable DHCP snooping globally.
   CONFIGURATION mode

   ```
   ip dhcp snooping
   ```
2. Specify ports connected to DHCP servers as trusted.
   INTERFACE mode

   ```
   ip dhcp snooping trust
   ```

3. Enable DHCP snooping on a VLAN.
   CONFIGURATION mode

   ```
   ip dhcp snooping vlan name
   ```

Adding a Static Entry in the Binding Table

To add a static entry in the binding table, use the following command.

- Add a static entry in the binding table.
  EXEC Privilege mode

  ```
  ip dhcp snooping binding mac
  ```

Clearing the Binding Table

To clear the binding table, use the following command.

- Delete all of the entries in the binding table.
  EXEC Privilege mode

  ```
  clear ip dhcp snooping binding
  ```

Displaying the Contents of the Binding Table

To display the contents of the binding table, use the following command.

- Display the contents of the binding table.
  EXEC Privilege mode

  ```
  show ip dhcp snooping
  ```

Example of the `show ip dhcp snooping` Command

View the DHCP snooping statistics with the `show ip dhcp snooping` command.

```
Dell#show ip dhcp snooping

IP DHCP Snooping                                 : Enabled.
IP DHCP Snooping Mac Verification                : Disabled.
IP DHCP Relay Information-option                 : Disabled.
IP DHCP Relay Trust Downstream                   : Disabled.

Database write-delay (In minutes)                : 0

DHCP packets information
Relay Information-option packets                 : 0
Relay Trust downstream packets                   : 0
Snooping packets                                 : 0

Packets received on snooping disabled L3 Ports   : 0
Snooping packets processed on L2 vlans           : 142

DHCP Binding File Details
Invalid File                                     : 0
Invalid Binding Entry                            : 0
Binding Entry lease expired                      : 0
List of Trust Ports                              :Te 0/49
List of DHCP Snooping Enabled Vlans              :Vl 10
List of DAI Trust ports                          :Te 0/49
```

Drop DHCP Packets on Snooped VLANs Only

Binding table entries are deleted when a lease expires or the relay agent encounters a DHCPRELEASE.

Line cards maintain a list of snooped VLANs. When the binding table fills, DHCP packets are dropped only on snooped VLANs, while such packets are forwarded across non-snooped VLANs. Because DHCP packets are dropped, no new IP address assignments are made. However, DHCP release and decline packets are allowed so that the DHCP snooping table can decrease in size. After the table usage falls below the maximum limit of 4000 entries, new IP address assignments are allowed.

To view the number of entries in the table, use the `show ip dhcp snooping binding` command. This output displays the snooping binding table created using the ACK packets from the trusted port.

```
Dell#show ip dhcp snooping binding

Codes : S - Static D - Dynamic

IP Address  MAC Address         Expires(Sec) Type VLAN   Interface
================================================================
10.1.1.251  00:00:4d:57:f2:50   172800       D    Vl 10   Te 0/2
10.1.1.252  00:00:4d:57:e6:f6   172800       D    Vl 10   Te 0/1
10.1.1.253  00:00:4d:57:f8:e8   172740       D    Vl 10   Te 0/3
10.1.1.254  00:00:4d:69:e8:f2   172740       D    Vl 10   Te 0/50

Total number of Entries in the table : 4
```

Dynamic ARP Inspection

Dynamic address resolution protocol (ARP) inspection prevents ARP spoofing by forwarding only ARP frames that have been validated against the DHCP binding table.

ARP is a stateless protocol that provides no authentication mechanism. Network devices accept ARP requests and replies from any device. ARP replies are accepted even when no request was sent. If a client receives an ARP message for which a relevant entry already exists in its ARP cache, it overwrites the existing entry with the new information.

The lack of authentication in ARP makes it vulnerable to spoofing. ARP spoofing is a technique attackers use to inject false IP-to-MAC mappings into the ARP cache of a network device. It is used to launch man-in-the-middle (MITM), and denial-of-service (DoS) attacks, among others.

A spoofed ARP message is one in which the MAC address in the sender hardware address field and the IP address in the sender protocol field are strategically chosen by the attacker. For example, in an MITM attack, the attacker sends a client an ARP message containing the attacker's MAC address and the gateway's IP address. The client then thinks that the attacker is the gateway, and sends all internet-bound packets to it. Likewise, the attacker sends the gateway an ARP message containing the attacker's MAC address and the client's IP address. The gateway then thinks that the attacker is the client and forwards all packets addressed to the client to it. As a result, the attacker is able to sniff all packets to and from the client.

Other attacks using ARP spoofing include:

Broadcast	An attacker can broadcast an ARP reply that specifies FF:FF:FF:FF:FF:FF as the gateway's MAC address, resulting in all clients broadcasting all internet-bound packets.

MAC flooding	An attacker can send fraudulent ARP messages to the gateway until the ARP cache is exhausted, after which, traffic from the gateway is broadcast.
Denial of service	An attacker can send a fraudulent ARP messages to a client to associate a false MAC address with the gateway address, which would blackhole all internet-bound packets from the client.

NOTE: Dynamic ARP inspection (DAI) uses entries in the L2SysFlow CAM region, a sub-region of SystemFlow. One CAM entry is required for every DAI-enabled VLAN. You can enable DAI on up to 16 VLANs on a system. However, the ExaScale default CAM profile allocates only nine entries to the L2SysFlow region for DAI. You can configure 10 to 16 DAI-enabled VLANs by allocating more CAM space to the L2SysFlow region before enabling DAI.

SystemFlow has 102 entries by default. This region is comprised of two sub-regions: L2Protocol and L2SystemFlow. L2Protocol has 87 entries; L2SystemFlow has 15 entries. Six L2SystemFlow entries are used by Layer 2 protocols, leaving nine for DAI. L2Protocol can have a maximum of 100 entries; you must expand this region to capacity before you can increase the size of L2SystemFlow. This is relevant when you are enabling DAI on VLANs. If, for example, you want to enable DAI on 16 VLANs, you need seven more entries; in this case, reconfigure the SystemFlow region for 122 entries using the `layer-2 eg-acl value fib value frrp value ing-acl value learn value l2pt value qos value system-flow 122` command.

The logic is as follows:

L2Protocol has 87 entries by default and must be expanded to its maximum capacity, 100 entries, before L2SystemFlow can be increased; therefore, 13 more L2Protocol entries are required. L2SystemFlow has 15 entries by default, but only nine are for DAI; to enable DAI on 16 VLANs, seven more entries are required. 87 L2Protocol + 13 additional L2Protocol + 15 L2SystemFlow + 7 additional L2SystemFlow equals 122.

Configuring Dynamic ARP Inspection

To enable dynamic ARP inspection, use the following commands.

1. Enable DHCP snooping.
2. Validate ARP frames against the DHCP snooping binding table.
   INTERFACE VLAN mode

   ```
   arp inspection
   ```

Examples of Viewing the ARP Database and Packets

To view entries in the ARP database, use the `show arp inspection database` command.

```
Dell#show arp inspection database

Protocol  Address     Age(min) Hardware Address   Interface VLAN   CPU
---------------------------------------------------------------------
Internet  10.1.1.251  -        00:00:4d:57:f2:50  Te 0/2    Vl 10  CP
Internet  10.1.1.252  -        00:00:4d:57:e6:f6  Te 0/1    Vl 10  CP
Internet  10.1.1.253  -        00:00:4d:57:f8:e8  Te 0/3    Vl 10  CP
Internet  10.1.1.254  -        00:00:4d:69:e8:f2  Te 0/50   Vl 10  CP
Dell#
```

To see how many valid and invalid ARP packets have been processed, use the `show arp inspection statistics` command.

```
Dell#show arp inspection statistics

Dynamic ARP Inspection (DAI) Statistics
---------------------------------------
Valid ARP Requests            : 0
Valid ARP Replies             : 1000
Invalid ARP Requests          : 1000
Invalid ARP Replies           : 0
Dell#
```

Bypassing the ARP Inspection

You can configure a port to skip ARP inspection by defining the interface as trusted, which is useful in multi-switch environments.
ARPs received on trusted ports bypass validation against the binding table. All ports are untrusted by default.

To bypass the ARP inspection, use the following command.

- Specify an interface as trusted so that ARPs are not validated against the binding table.

  INTERFACE mode

  ```
  arp inspection-trust
  ```

Dell Networking OS Behavior: Introduced in Dell Networking OS version 8.2.1.0, DAI was available for Layer 3 only. However, Dell Networking OS version 8.2.1.1 extends DAI to Layer 2.

Source Address Validation

Using the DHCP binding table, Dell Networking OS can perform three types of source address validation (SAV).

Table 19. Three Types of Source Address Validation

Source Address Validation	Description
IP Source Address Validation	Prevents IP spoofing by forwarding only IP packets that have been validated against the DHCP binding table.
DHCP MAC Source Address Validation	Verifies a DHCP packet's source hardware address matches the client hardware address field (CHADDR) in the payload.
IP+MAC Source Address Validation	Verifies that the IP source address and MAC source address are a legitimate pair.

Enabling IP Source Address Validation

IP source address validation (SAV) prevents IP spoofing by forwarding only IP packets that have been validated against the DHCP binding table.
A spoofed IP packet is one in which the IP source address is strategically chosen to disguise the attacker. For example, using ARP spoofing, an attacker can assume a legitimate client's identity and receive traffic addressed to it. Then the attacker can spoof the client's IP address to interact with other clients.

The DHCP binding table associates addresses the DHCP servers assign, with the port on which the requesting client is attached. When you enable IP source address validation on a port, the system verifies that the source IP address is one that is associated with the incoming port. If an attacker is impostering as a legitimate client, the source address appears on the wrong ingress port and the system drops the packet. Likewise, if the IP address is fake, the address is not on the list of permissible addresses for the port and the packet is dropped.

To enable IP source address validation, use the following command.

NOTE: If you enable IP source guard using the `ip dhcp source-address-validation` command and there are 187 entries or more in the current DHCP snooping binding table, SAV may not be applied to all entries. To ensure that SAV is applied correctly to all entries, enable the `ip dhcp source-address-validation` command before adding entries to the binding table.

- Enable IP source address validation.

  INTERFACE mode

  ```
  ip dhcp source-address-validation
  ```

DHCP MAC Source Address Validation

DHCP MAC source address validation (SAV) validates a DHCP packet's source hardware address against the client hardware address field (CHADDR) in the payload.
Dell Networking OS ensures that the packet's source MAC address is checked against the CHADDR field in the DHCP header only for packets from snooped VLANs.

- Enable DHCP MAC SAV.

  CONFIGURATION mode

  ```
  ip dhcp snooping verify mac-address
  ```

Enabling IP+MAC Source Address Validation

The following feature is available on the S6000 platform.
IP source address validation (SAV) validates the IP source address of an incoming packet against the DHCP snooping binding table. IP+MAC SAV ensures that the IP source address and MAC source address are a legitimate pair, rather than validating each attribute individually. You cannot configure IP+MAC SAV with IP SAV.

1. Allocate at least one FP block to the ipmacacl CAM region.

   CONFIGURATION mode

   ```
   cam-acl l2acl
   ```

2. Save the running-config to the startup-config.

   EXEC Privilege mode

   ```
   copy running-config startup-config
   ```

3. Reload the system.

   EXEC Privilege

   ```
   reload
   ```

4. Enable IP+MAC SAV.

   INTERFACE mode

```
ip dhcp source-address-validation ipmac
```

Dell Networking OS creates an ACL entry for each IP+MAC address pair in the binding table and applies it to the interface.

To display the IP+MAC ACL for an interface for the entire system, use the `show ip dhcp snooping source-address-validation [`*`interface`*`]` command in EXEC Privilege mode.

13

Equal Cost Multi-Path (ECMP)

Equal cost multi-path (ECMP) is supported on theS6000 platform.

ECMP for Flow-Based Affinity

ECMP for flow-based affinity is available on theplatform.

Flow-based affinity includes the following:

- Link Bundle Monitoring

Configuring the Hash Algorithm

TeraScale has one algorithm that is used for link aggregation groups (LAGs), ECMP, and NH-ECMP, and ExaScale can use three different algorithms for each of these features.
To adjust the ExaScale behavior to match TeraScale, use the following command.

- Change the ExaScale hash-algorithm for LAG, ECMP, and NH-ECMP to match TeraScale.

  CONFIGURATION mode.

  ```
  hash-algorithm ecmp checksum 0 lag checksum 0 nh-ecmp checksum 0
  ```

Dell Networking OS Behavior: In the Dell Networking OS versions prior to 8.2.1.2, the ExaScale default hash-algorithm is 0. Beginning with Dell Networking OS version 8.2.1.2, the default hash-algorithm is 24.

Enabling Deterministic ECMP Next Hop

Deterministic ECMP next hop arranges all ECMPs in order before writing them into the content addressable memory (CAM).
For example, suppose the RTM learns eight ECMPs in the order that the protocols and interfaces came up. In this case, the forwarding information base (FIB) and CAM sort them so that the ECMPs are always arranged. This implementation ensures that every chassis having the same prefixes orders the ECMPs the same.

With eight or less ECMPs, the ordering is lexicographic and deterministic. With more than eight ECMPs, ordering is deterministic, but it is not in lexicographic order.

To enable deterministic ECMP next hop, use the appropriate command.

NOTE: Packet loss might occur when you enable `ip/ipv6 ecmp-deterministic` for the first-time only.

- Enable IPv4 Deterministic ECMP Next Hop.

  CONFIGURATION mode.

  ```
  ip ecmp-deterministic
  ```
- Enable IPv6 Deterministic ECMP Next Hop.

CONFIGURATION mode.

```
ipv6 ecmp-deterministic
```

Configuring the Hash Algorithm Seed

Deterministic ECMP sorts ECMPs in order even though RTM provides them in a random order. However, the hash algorithm uses as a seed the lower 12 bits of the chassis MAC, which yields a different hash result for every chassis.
This behavior means that for a given flow, even though the prefixes are sorted, two unrelated chassis can select different hops.

Dell Networking OS provides a command line interface (CLI)-based solution for modifying the hash seed to ensure that on each configured system, the ECMP selection is same. When configured, the same seed is set for ECMP, LAG, and NH, and is used for incoming traffic only.

NOTE: While the seed is stored separately on each port-pipe, the same seed is used across all CAMs.

NOTE: You cannot separate LAG and ECMP, but you can use different algorithms across the chassis with the same seed. If LAG member ports span multiple port-pipes and line cards, set the seed to the same value on each port-pipe to achieve deterministic behavior.

NOTE: If you remove the hash algorithm configuration, the hash seed does not return to the original factory default setting.

To configure the hash algorithm seed, use the following command.

- Specify the hash algorithm seed.
  CONFIGURATION mode.

  ```
  hash-algorithm seed value [stack—unit number] [port-set number]
  ```

  The range is from 0 to 4095.

Link Bundle Monitoring

Link bundle monitoring is supported on the S6000 platform.

Monitoring linked ECMP bundles allows traffic distribution amounts in a link to be monitored for unfair distribution at any given time. A default threshold of 60% is defined as an acceptable amount of traffic on a member link. Links are monitored in 15-second intervals for three consecutive instances. Any deviation within that time causes a syslog to be sent and an alarm event to be generated. When the deviation clears, another syslog is sent and a clear alarm event is generated. for example Link bundle monitoring percent threshold `%STKUNIT0-M:CP %IFMGR-5-BUNDLE_UNEVEN_DISTRIBUTION: Found uneven distribution in LAG bundle 11..`

The link bundle utilization is calculated as the total bandwidth of all links divided by the total bytes-per-second of all links. Within each ECMP group, interfaces can be specified. If monitoring is enabled for the ECMP group, the utilization calculation is performed when the utilization of the link-bundle (not a link within a bundle) exceeds 60%.

Enable link bundle monitoring using the **ecmp-group** command.

NOTE: An ecmp-group index is generated automatically for each unique ecmp-group when the user configures multipath routes to the same network. The system can generate a maximum of 512 unique ecmp-groups. The ecmp-group indexes are generated in even numbers (0, 2, 4, 6... 1022) and are for information only.

For link bundle monitoring with ECMP, the ecmp-group command is used to enable the link bundle monitoring feature. The ecmp-group with id 2, enabled for link bundle monitoring is user configured. This is different from the ecmp-group index 2 that is created by configuring routes and is automatically generated.

These two ecmp-groups are not related in any way.

Example of Viewing Link Bundle Monitoring

```
Dell# show link-bundle-distribution ecmp-group 1
Link-bundle trigger threshold - 60
ECMP bundle - 1 Utilization[In Percent] - 44 Alarm State - Active
Interface  Line Protocol  Utilization[In Percent]
Te 0/0     Up             36
Te 0/1     Up             52
```

Managing ECMP Group Paths

Managing ECMP group paths is supported only on the S6000 platform.
Configure the maximum number of paths for an ECMP route that the L3 CAM can hold to avoid path degeneration. When you do not configure the maximum number of routes, the CAM can hold a maximum ECMP per route.

To configure the maximum number of paths, use the following command.

NOTE: Save the new ECMP settings to the startup-config (`write-mem`) then reload the system for the new settings to take effect.

- Configure the maximum number of paths per ECMP group.

  CONFIGURATION mode.

  ```
  ip ecmp-group maximum-paths {2-64}
  ```
- Enable ECMP group path management.

  CONFIGURATION mode.

  ```
  ip ecmp-group path-fallback
  ```

Example of the `ip ecmp-group maximum-paths` Command

```
Dell(conf)#ip ecmp-group maximum-paths 3
User configuration has been changed. Save the configuration and reload to take
effect
Dell(conf)#
```

Creating an ECMP Group Bundle

Within each ECMP group, you can specify an interface.
If you enable monitoring for the ECMP group, the utilization calculation is performed when the average utilization of the link-bundle (as opposed to a single link within the bundle) exceeds 60%.

1. Create a user-defined ECMP group bundle.
   CONFIGURATION mode

   ```
   ecmp-group ecmp-group-id
   ```

   The range is from 1 to 64.
2. Add interfaces to the ECMP group bundle.
   CONFIGURATION ECMP-GROUP mode

   ```
   interface interface interface tengigabitethernet 0/0 interface port-channel
   100
   ```
3. Enable the monitoring for the bundle.
   CONFIGURATION ECMP-GROUP mode

   ```
   link-bundle-monitor enable
   ```

Modifying the ECMP Group Threshold

You can customize the threshold percentage for monitoring ECMP group bundles.
To customize the ECMP group bundle threshold and to view the changes, use the following commands.

- Modify the threshold for monitoring ECMP group bundles.
  CONFIGURATION mode

  ```
  link-bundle-distribution trigger-threshold {percent}
  ```

  The range is from 1 to 90%.

  The default is **60%**.
- Display details for an ECMP group bundle.
  EXEC mode

  ```
  show link-bundle-distribution ecmp-group ecmp-group-id
  ```

  The range is from 1 to 64.

Viewing an ECMP Group

NOTE: An ecmp-group index is generated automatically for each unique ecmp-group when you configure multipath routes to the same network. The system can generate a maximum of 512 unique ecmp-groups. The ecmp-group indices are generated in even numbers (0, 2, 4, 6... 1022) and are for information only.

You can configure ecmp-group with *id 2* for link bundle monitoring. This ecmp-group is different from the ecmp-group *index 2* that is created by configuring routes and is automatically generated. These two ecmp-groups are not related in any way.

```
Dell(conf-ecmp-group-5)#show config
!
ecmp-group 5
  interface tengigabitethernet 0/2
  interface tengigabitethernet 0/3
  link-bundle-monitor enable
Dell(conf-ecmp-group-5)#
```

Support for /128 IPv6 and /32 IPv4 Prefixes in Layer 3 Host Table and LPM Table

IPv6 enhancements utilize the capability on S6000 platform to program /128 IPv6 prefixes in LPM table and /32 IPv4 prefixes in Host table. Also host table provides ECMP support for destination prefixes in the hardware. S6000 platform uses the hardware chip that supports this behavior and hence they can make use of this capability.

CLI commands are introduced to move /128 IPv6 prefix route entries and /32 IPv4 prefix route entries from Host table to LPM table and vice versa. When moving the destination prefixes from Route to Host table, there is a possibility of getting into hash collision because the Host table on the device is a Hash table. In this scenario, a workaround does not exist for the user having route entries programmed in host table.

When the command is issued, you are prompted with a warning message stating that the command configuration can take effect on existing prefixes only when “clear ip route *” command is used. When you use the clear command, all the existing /32 IPv4 prefix route entries are reprogrammed in appropriate table. Also, all the other existing IPv4 entries are removed and reprogrammed as a result of the clear command.

Dell Networking OS releases earlier than Release 9.3(0.1) stores IPv6 /128 entries in Host table since it cannot be written in LPM table, and IPv4 0/32 route entries are written in LPM table itself to support the ECMP since ECMP was not supported in Host table. On S6000 switches, unified forwarding table (UFT) is enabled, and the host table size is bigger compared to the LPM. When you move the IPv4 /32 route prefix entry in host table, more space is obtained that can be utilized for other route prefix entries.

Support for ECMP in host table

ECMP support in the L3 host table is available only on S6000 platform. IPv6 /128 prefix route entries and IPv4 /32 prefix entries which are moved to host table can have ECMP. For other platforms, only the IPv6 /128 prefix route entries is stored in the L3 host table without ECMP support.

The software supports a command to program IPv6 /128 route prefixes in the host table.

The output of show IPv6 cam command has been enhanced to include the ECMP field in the Neighbor table of Ipv6 CAM. The sample output is displayed as follows, which is similar to the prefix table.

The following is the portion of the example output from Release 9.3(0.1):

```
Neighbor                                          Mac-Addr
Port      VId
------------------------------------           -----------------
---------    ----
[  132] 20::1                                     00:00:20:d5:ec:a0      Fo
0/116          0
```

The following is the portion of the example output from Release 9.3(0.1):

```
Neighbor Mac-Addr                           Port          Vid     EC
-------------------------------------           -------------------------
[  132] 20::1 00:00:20:d5:ec:a0            Fo 0/116      0        1
[  132] 20::1 00:00:20:d5:ec:a1            Fo 0/114      0        1
```

Support for moving /128 IPv6 Prefixes and /32 IPv4 Prefixes

The software supports a command to program IPv6 /128 route prefixes in the route table. You can define IPv6 /128 route prefixes in the route table using the `ipv6 unicast-host-route`command. You can also define IPv4 /32 route prefixes in the host table using the `ipv4 unicast-host-route`command.

14

FCoE Transit

The Fibre Channel over Ethernet (FCoE) Transit feature is supported on the switch on Ethernet interfaces. When you enable the switch for FCoE transit, the switch functions as a FIP snooping bridge.

NOTE: FIP snooping is not supported on Fibre Channel interfaces or in a switch stack.

Fibre Channel over Ethernet

FCoE provides a converged Ethernet network that allows the combination of storage-area network (SAN) and LAN traffic on a Layer 2 link by encapsulating Fibre Channel data into Ethernet frames.

FCoE works with the Ethernet enhancements provided in data center bridging (DCB) to support lossless (no-drop) SAN and LAN traffic. In addition, DCB provides flexible bandwidth sharing for different traffic types, such as LAN and SAN, according to 802.1p priority classes of service. DCBx should be enabled on the system before the FIP snooping feature is enabled. For more information, refer to the Data Center Bridging (DCB) chapter.

Ensure Robustness in a Converged Ethernet Network

Fibre Channel networks used for SAN traffic employ switches that operate as trusted devices. To communicate with other end devices attached to the Fibre Channel network, end devices log into the switch to which they are attached.

Because Fibre Channel links are point-to-point, a Fibre Channel switch controls all storage traffic that an end device sends and receives over the network. As a result, the switch can enforce zoning configurations, ensure that end devices use their assigned addresses, and secure the network from unauthorized access and denial-of-service (DoS) attacks.

To ensure similar Fibre Channel robustness and security with FCoE in an Ethernet cloud network, FIP establishes virtual point-to-point links between FCoE end-devices (server ENodes and target storage devices) and FCoE forwarders (FCFs) over transit FCoE-enabled bridges.

Ethernet bridges commonly provide ACLs that can emulate a point-to-point link by providing the traffic enforcement required to create a Fibre Channel-level of robustness. You can configure ACLs to emulate point-to-point links, providing control over the traffic received or transmitted into the switch. To automatically generate ACLs, use FIP snooping. In addition, FIP serves as a Layer 2 protocol to:

- Operate between FCoE end-devices and FCFs over intermediate Ethernet bridges to prevent unauthorized access to the network and achieve the required security.
- Allow transit Ethernet bridges to efficiently monitor FIP frames passing between FCoE end-devices and an FCF. To dynamically configure ACLs on the bridge to only permit traffic authorized by the FCF, use the FIP snooping data.

FIP enables FCoE devices to discover one another, initialize and maintain virtual links over an Ethernet network, and access storage devices in a storage area network (SAN). FIP satisfies the Fibre Channel requirement for point-to-point connections by creating a unique virtual link for each connection between an FCoE end-device and an FCF via a transit switch.

FIP provides functionality for discovering and logging into an FCF. After discovering and logging in, FIP allows FCoE traffic to be sent and received between FCoE end-devices (ENodes) and the FCF. FIP uses its own EtherType and frame format. The following illustration shows the communication that occurs between an ENode server and an FCoE switch (FCF).

The following table lists the FIP functions.

Table 20. FIP Functions

FIP Function	Description
FIP VLAN discovery	FCoE devices (ENodes) discover the FCoE VLANs on which to transmit and receive FIP and FCoE traffic.
FIP discovery	FCoE end-devices and FCFs are automatically discovered.
Initialization	FCoE devices learn ENodes from the FLOGI and FDISC to allow immediate login and create a virtual link with an FCoE switch.
Maintenance	A valid virtual link between an FCoE device and an FCoE switch is maintained and the LOGO functions properly.
Logout	On receiving a FLOGI packet, FSB deletes all existing sessions from the ENode to the FCF.

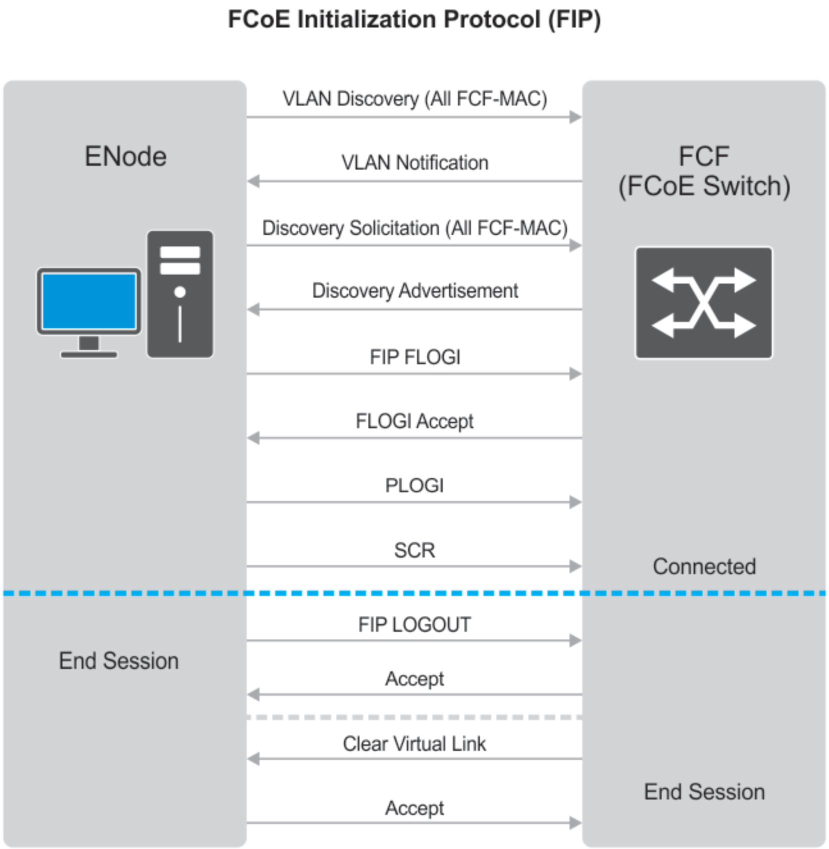

FCoE Initialization Protocol (FIP)
ENode
FCF
(FCoE Switch)
VLAN Discovery (All FCF-MAC)
VLAN Notification
Discovery Solicitation (All FCF-MAC)
Discovery Advertisement
FIP FLOGI
FLOGI Accept
PLOGI
SCR
Connected
End Session
FIP LOGOUT
Accept
Clear Virtual Link
Accept
End Session

Figure 33. FIP Discovery and Login Between an ENode and an FCF

FIP Snooping on Ethernet Bridges

In a converged Ethernet network, intermediate Ethernet bridges can snoop on FIP packets during the login process on an FCF. Then, using ACLs, a transit bridge can permit only authorized FCoE traffic to be transmitted between an FCoE end-device and an FCF. An Ethernet bridge that provides these functions is called a FIP snooping bridge (FSB).

On a FIP snooping bridge, ACLs are created dynamically as FIP login frames are processed. The ACLs are installed on switch ports configured for ENode mode for server-facing ports and FCF mode for a trusted port directly connected to an FCF.

Enable FIP snooping on the switch, configure the FIP snooping parameters, and configure CAM allocation for FCoE. When you enable FIP snooping, all ports on the switch by default become ENode ports.

Dynamic ACL generation on the switch operating as a FIP snooping bridge function as follows:

Port-based ACLs	These ACLs are applied on all three port modes: on ports directly connected to an FCF, server-facing ENode ports, and bridge-to-bridge links. Port-based ACLs take precedence over global ACLs.

FCoE-generated ACLs	These take precedence over user-configured ACLs. A user-configured ACL entry cannot deny FCoE and FIP snooping frames.

The following illustration shows a switch used as a FIP snooping bridge in a converged Ethernet network. The top-of-rack (ToR) switch operates as an FCF for FCoE traffic. Converged LAN and SAN traffic is transmitted between the ToR switch and an switch. The switch operates as a lossless FIP snooping bridge to transparently forward FCoE frames between the ENode servers and the FCF switch.

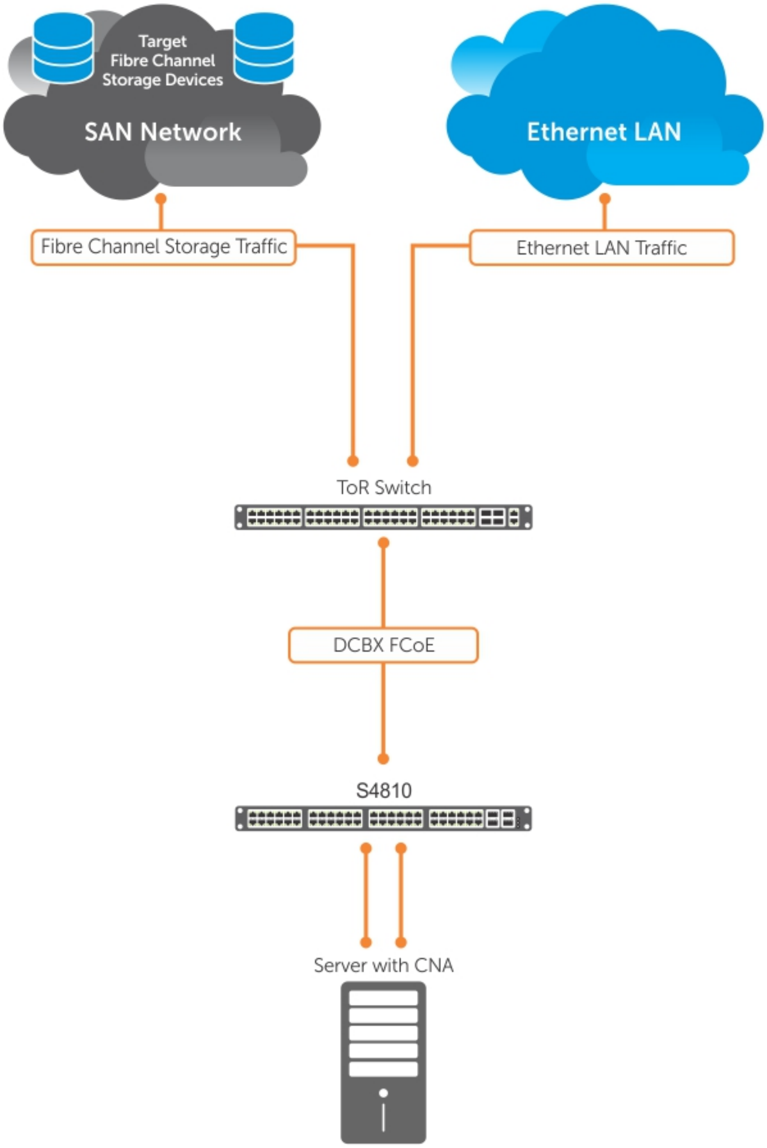

Target
Fibre Channel
Storage Devices
SAN Network
Ethernet LAN
Fibre Channel Storage Traffic
Ethernet LAN Traffic
ToR Switch
DCBX FCoE
S4810
Server with CNA

Figure 34. FIP Snooping on an S4810 Switch

The following sections describe how to configure the FIP snooping feature on a switch that functions as a FIP snooping bridge so that it can perform the following functions:

- Allocate CAM resources for FCoE.
- Perform FIP snooping (allowing and parsing FIP frames) globally on all VLANs or on a per-VLAN basis.
- To assign a MAC address to an FCoE end-device (server ENode or storage device) after a server successfully logs in, set the FCoE MAC address prefix (FC-MAP) value an FCF uses. The FC-MAP value is used in the ACLs installed in bridge-to-bridge links on the switch.
- To provide more port security on ports that are directly connected to an FCF and have links to other FIP snooping bridges, set the FCF or Bridge-to-Bridge Port modes.
- To ensure that they are operationally active, check FIP snooping-enabled VLANs.
- Process FIP VLAN discovery requests and responses, advertisements, solicitations, FLOGI/FDISC requests and responses, FLOGO requests and responses, keep-alive packets, and clear virtual-link messages.

FIP Snooping in a Switch Stack

FIP snooping supports switch stacking as follows:

- A switch stack configuration is synchronized with the standby stack unit.
- Dynamic population of the FCoE database (ENode, Session, and FCF tables) is synchronized with the standby stack unit. The FCoE database is maintained by snooping FIP keep-alive messages.
- In case of a failover, the new master switch starts the required timers for the FCoE database tables. Timers run only on the master stack unit.

Using FIP Snooping

There are four steps to configure FCoE transit.

1. Enable the FCoE transit feature on a switch to maintain FIP snooping information on the switch.
2. Enable FIP snooping globally on all Virtual Local Area Networks (VLANs) or individual VLANs on a FIP snooping bridge.
3. Configure the FC-Map value applied globally by the switch on all VLANs or an individual VLAN.
4. Configure FCF mode for a FIP snooping bridge-to-FCF link.

For a sample FIP snooping configuration, refer to FIP Snooping Configuration Example.

Statisical information is available for FIP Snooping-related information. For available commands, refer to the *FCoE Transit* chapter in the *Dell Networking OS Command Line Reference Guide*.

FIP Snooping Prerequisites

Before you enable FCoE transit and configure FIP snooping on a switch, ensure that certain conditions are met.

A FIP snooping bridge requires data center bridging exchange protocol (DCBx) and priority-based flow control (PFC) to be enabled on the switch for lossless Ethernet connections (refer to the Data Center Bridging (DCB) chapter). Dell Networking recommends also enabling enhanced transmission selection (ETS); however, ETS is recommended but not required.

If you enable DCBx and PFC mode is on (PFC is operationally up) in a port configuration, FIP snooping is operational on the port. If the PFC parameters in a DCBx exchange with a peer are not synchronized, FIP and FCoE frames are dropped on the port after you enable the FIP snooping feature.

For VLAN membership, you must:

- create the VLANs on the switch which handles FCoE traffic (use the `interface vlan` command).
- configure each FIP snooping port to operate in Hybrid mode so that it accepts both tagged and untagged VLAN frames (use the `portmode hybrid` command).
- configure tagged VLAN membership on each FIP snooping port that sends and receives FCoE traffic and has links with an FCF, ENode server, or another FIP snooping bridge (use the `tagged port-type slot/port` command).

The default VLAN membership of the port must continue to operate with untagged frames. FIP snooping is not supported on a port that is configured for non-default untagged VLAN membership.

Important Points to Remember

- Enable DCBx on the switch before enabling the FIP Snooping feature.
- To enable the feature on the switch, configure FIP Snooping.
- To allow FIP frames to pass through the switch on all VLANs, enable FIP snooping globally on a switch.
- A switch can support a maximum eight VLANs. Configure at least one FCF/bridge-to-bridge port mode interface for any FIP snooping-enabled VLAN.
- You can configure multiple FCF-trusted interfaces in a VLAN.
- When you disable FIP snooping:
  - ACLs are not installed, FIP and FCoE traffic is not blocked, and FIP packets are not processed.
  - The existing per-VLAN and FIP snooping configuration is stored. The configuration is re-applied the next time you enable the FIP snooping feature.
- You must apply the CAM-ACL space for the FCoE region before enabling the FIP-Snooping feature. If you do not apply CAM-ACL space the following error message is displayed:

```
Dell(conf)#feature fip-snooping
% Error: Cannot enable fip snooping. CAM Region not allocated for Fcoe.
Dell(conf)#
```

NOTE: You must manually add the CAM-ACL space to the FCoE region, as it is not applied by default.

Enabling the FCoE Transit Feature

The following sections describe how to enable FCoE transit.

NOTE: FCoE transit is disabled by default. To enable this feature, you must follow the Configuring FIP Snooping.

As soon as you enable the FCoE transit feature on a switch-bridge, existing VLAN-specific and FIP snooping configurations are applied. The FCoE database is populated when the switch connects to a converged network adapter (CNA) or FCF port and compatible DCB configurations are synchronized. By default, all FCoE and FIP frames are dropped unless specifically permitted by existing FIP snooping-generated ACLs. You can reconfigure any of the FIP snooping settings.

If you disable FCoE transit, FIP and FCoE traffic are handled as normal Ethernet frames and no FIP snooping ACLs are generated. The VLAN-specific and FIP snooping configuration is disabled and stored until you re-enable FCoE transit and the configurations are re-applied.

Enable FIP Snooping on VLANs

You can enable FIP snooping globally on a switch on all VLANs or on a specified VLAN.

When you enable FIP snooping on VLANs:

- FIP frames are allowed to pass through the switch on the enabled VLANs and are processed to generate FIP snooping ACLs.
- FCoE traffic is allowed on VLANs only after a successful virtual-link initialization (fabric login FLOGI) between an ENode and an FCF. All other FCoE traffic is dropped.
- You must configure at least one interface for FCF (FIP snooping bridge-bridge) mode on a FIP snooping-enabled VLAN. You can configure multiple FCF trusted interfaces in a VLAN.
- A maximum of eight VLANS are supported for FIP snooping on the switch. When enabled globally, FIP snooping processes FIP packets in traffic only from the first eight incoming VLANs. When enabled on a per-VLAN basis, FIP snooping is supported on up to eight VLANs.

Configure the FC-MAP Value

You can configure the FC-MAP value to be applied globally by the switch on all or individual FCoE VLANs to authorize FCoE traffic.

The configured FC-MAP value is used to check the FC-MAP value for the MAC address assigned to ENodes in incoming FCoE frames. If the FC-MAP value does not match, FCoE frames are dropped. A session between an ENode and an FCF is established by the switch-bridge only when the FC-MAP value on the FCF matches the FC-MAP value on the FIP snooping bridge.

Configure a Port for a Bridge-to-Bridge Link

If a switch port is connected to another FIP snooping bridge, configure the FCoE-Trusted Port mode for bridge-bridge links.

Initially, all FCoE traffic is blocked. Only FIP frames with the ALL_FCF_MAC and ALL_ENODE_MAC values in their headers are allowed to pass. After the switch learns the MAC address of a connected FCF, it allows FIP frames destined to or received from the FCF MAC address.

FCoE traffic is allowed on the port only after the switch learns the FC-MAP value associated with the specified FCF MAC address and verifies that it matches the configured FC-MAP value for the FCoE VLAN.

Configure a Port for a Bridge-to-FCF Link

If a port is directly connected to an FCF, configure the port mode as FCF. Initially, all FCoE traffic is blocked; only FIP frames are allowed to pass.

FCoE traffic is allowed on the port only after a successful fabric login (FLOGI) request/response and confirmed use of the configured FC-MAP value for the VLAN.

FLOGI and fabric discovery (FDISC) request/response packets are trapped to the CPU. They are forwarded after the necessary ACLs are installed.

Impact on Other Software Features

When you enable FIP snooping on a switch, other software features are impacted. The following table lists the impact of FIP snooping.

Table 21. Impact of Enabling FIP Snooping

Impact	Description
MAC address learning	MAC address learning is not performed on FIP and FCoE frames, which are denied by ACLs dynamically created by FIP snooping on server-facing ports in ENode mode.
MTU auto-configuration	MTU size is set to mini-jumbo (2500 bytes) when a port is in Switchport mode, the FIP snooping feature is enabled on the switch, and FIP snooping is enabled on all or individual VLANs.
Link aggregation group (LAG)	FIP snooping is supported on port channels on ports on which PFC mode is on (PFC is operationally up).
STP	If you enable an STP protocol (STP, RSTP, PVSTP, or MSTP) on the switch and ports enter a blocking state, when the state change occurs, the corresponding port-based ACLs are deleted. If a port is enabled for FIP snooping in ENode or FCF mode, the ENode/FCF MAC-based ACLs are deleted.

FIP Snooping Restrictions

The following restrictions apply when you configure FIP snooping.

- The maximum number of FCoE VLANs supported on the switch is eight.
- The maximum number of FIP snooping sessions supported per ENode server is 32. To increase the maximum number of sessions to 64, use the `fip-snooping max-sessions-per-enodemac` command.
- The maximum number of FCFs supported per FIP snooping-enabled VLAN is twelve.

- The maximum number of FCoE VLANs supported on the switch is eight.
- The maximum number of FIP snooping sessions (including NPIV sessions) supported per ENode server is 16
- In a full FCoE N_port ID virtualization (NPIV) configuration, 16 sessions (one FLOGI + fifteen NPIV sessions) are supported per ENode. In an FCoE NPIV configuration, only one session is supported per ENode.
- The maximum number of FCFs supported per FIP snooping-enabled VLAN is four.
- Links to other FIP snooping bridges on a FIP snooping-enabled port (bridge-to-bridge links) are not supported on the S6000 Switch.

Configuring FIP Snooping

You can enable FIP snooping globally on all FCoE VLANs on a switch or on an individual FCoE VLAN. By default, FIP snooping is disabled.

To enable FCoE transit on the switch and configure the FCoE transit parameters on ports, follow these steps.

1. Configure FCoE.

FCoE configuration:

```
copy flash:/ CONFIG_TEMPLATE/ FCoE_DCB_Config running-config
```

The configuration files are stored in the flash memory in the CONFIG_TEMPLATE file.

NOTE: DCB/DCBx is enabled when either of these configurations is applied.

2. Save the configuration on the switch.
   EXEC Privilege mode.

   ```
   write memory
   ```

3. Reload the switch to enable the configuration.
   EXEC Privilege mode.

   ```
   reload
   ```

   After the switch is reloaded, DCB/DCBx is enabled.

4. Enable the FCoE transit feature on a switch.
   CONFIGURATION mode.

   ```
   feature fip-snooping
   ```

5. Enable FIP snooping on all VLANs or on a specified VLAN.
   CONFIGURATION mode or VLAN INTERFACE mode.

   ```
   fip-snooping enable
   ```

6. Configure the port for bridge-to-FCF links.
   INTERFACE mode or CONFIGURATION mode

   ```
   fip-snooping port-mode fcf
   ```

NOTE: To disable the FCoE transit feature or FIP snooping on VLANs, use the `no` version of a command; for example, `no feature fip-snooping` or `no fip-snooping enable`.

Displaying FIP Snooping Information

Use the following `show` commands to display information on FIP snooping, .

Table 22. Displaying FIP Snooping Information

Command	Output
`show fip-snooping sessions [interface vlan vlan-id]`	Displays information on FIP-snooped sessions on all VLANs or a specified VLAN, including the ENode interface and MAC address, the FCF interface and MAC address, VLAN ID, FCoE MAC address and FCoE session ID number (FC-ID), worldwide node name (WWNN) and the worldwide port name (WWPN).
`show fip-snooping config`	Displays the FIP snooping status and configured FC-MAP values.
`show fip-snooping enode [enode-mac-address]`	Displays information on the ENodes in FIP-snooped sessions, including the ENode interface

Command	Output
	and MAC address, FCF MAC address, VLAN ID and FC-ID.
`show fip-snooping fcf [`*`fcf-mac-address`*`]`	Displays information on the FCFs in FIP-snooped sessions, including the FCF interface and MAC address, FCF interface, VLAN ID, FC-MAP value, FKA advertisement period, and number of ENodes connected.
`clear fip-snooping database interface vlan` *`vlan-id`* `{`*`fcoe-mac-address`* `|` *`enode-mac-address`* `|` *`fcf-mac-address`*`}`	Clears FIP snooping information on a VLAN for a specified FCoE MAC address, ENode MAC address, or FCF MAC address, and removes the corresponding ACLs generated by FIP snooping.
`show fip-snooping statistics [interface vlan` *`vlan-id`*`| interface` *`port-type port/slot`* `| interface port-channel` *`port-channel-number`*`]`	Displays statistics on the FIP packets snooped on all interfaces, including VLANs, physical ports, and port channels.
`clear fip-snooping statistics [interface vlan` *`vlan-id`* `| interface` *`port-type port/slot`* `| interface port-channel` *`port-channel-number`*`]`	Clears the statistics on the FIP packets snooped on all VLANs, a specified VLAN, or a specified port interface.
`show fip-snooping system`	Displays information on the status of FIP snooping on the switch (enabled or disabled), including the number of FCoE VLANs, FCFs, ENodes, and currently active sessions.
`show fip-snooping vlan`	Displays information on the FCoE VLANs on which FIP snooping is enabled.

Examples of the `show fip-snooping` Commands

The following example shows the `show fip-snooping sessions` command.

```
Dell#show fip-snooping sessions
Enode MAC          Enode Intf  FCF MAC            FCF Intf  VLAN
aa:bb:cc:00:00:00  Te 0/42     aa:bb:cd:00:00:00  Te 0/43   100
aa:bb:cc:00:00:00  Te 0/42     aa:bb:cd:00:00:00  Te 0/43   100
aa:bb:cc:00:00:00  Te 0/42     aa:bb:cd:00:00:00  Te 0/43   100
aa:bb:cc:00:00:00  Te 0/42     aa:bb:cd:00:00:00  Te 0/43   100
aa:bb:cc:00:00:00  Te 0/42     aa:bb:cd:00:00:00  Te 0/43   100

FCoE MAC          FC-ID         Port WWPN                Port WWNN
0e:fc:00:01:00:01 01:00:01 31:00:0e:fc:00:00:00:00 21:00:0e:fc:00:00:00:00
0e:fc:00:01:00:02 01:00:02 41:00:0e:fc:00:00:00:00 21:00:0e:fc:00:00:00:00
0e:fc:00:01:00:03 01:00:03 41:00:0e:fc:00:00:00:01 21:00:0e:fc:00:00:00:00
0e:fc:00:01:00:04 01:00:04 41:00:0e:fc:00:00:00:02 21:00:0e:fc:00:00:00:00
0e:fc:00:01:00:05 01:00:05 41:00:0e:fc:00:00:00:03 21:00:0e:fc:00:00:00:00
```

The following table describes the `show fip-snooping sessions` command fields.

Table 23. `show fip-snooping sessions` Command Description

Field	Description
ENode MAC	MAC address of the ENode .
ENode Interface	Slot/ port number of the interface connected to the ENode.
FCF MAC	MAC address of the FCF.
FCF Interface	Slot/ port number of the interface to which the FCF is connected.
VLAN	VLAN ID number used by the session.
FCoE MAC	MAC address of the FCoE session assigned by the FCF.
FC-ID	Fibre Channel ID assigned by the FCF.
Port WWPN	Worldwide port name of the CNA port.
Port WWNN	Worldwide node name of the CNA port.

The following example shows the `show fip-snooping config` command.

```
Dell# show fip-snooping config
FIP Snooping Feature enabled Status: Enabled
FIP Snooping Global enabled Status: Enabled
Global FC-MAP Value: 0X0EFC00

FIP Snooping enabled VLANs
VLAN   Enabled  FC-MAP
----  -------   --------
100   TRUE      0X0EFC00
```

The following example shows the `show fip-snooping enode` command.

```
Dell# show fip-snooping enode
Enode MAC          Enode Interface FCF MAC          VLAN  FC-ID
---------          --------------- -------          ----  -----
d4:ae:52:1b:e3:cd  Te 0/11        54:7f:ee:37:34:40 100   62:00:11
```

The following table describes the `show fip-snooping enode` command fields.

Table 24. `show fip-snooping enode` Command Description

Field	Description
ENode MAC	MAC address of the ENode.
ENode Interface	Slot/ port number of the interface connected to the ENode.
FCF MAC	MAC address of the FCF.
VLAN	VLAN ID number used by the session.
FC-ID	Fibre Channel session ID assigned by the FCF.

The following example shows the `show fip-snooping fcf` command.

```
Dell# show fip-snooping fcf
FCF MAC           FCF Interface VLAN FC-MAP   FKA_ADV_PERIOD No. of Enodes
-------           ------------- ---- ------   -------------- -------------
54:7f:ee:37:34:40 Po 22         100  0e:fc:00 4000           2
```

The following table describes the `show fip-snooping fcf` command fields.

Table 25. `show fip-snooping fcf` Command Description

Field	Description
FCF MAC	MAC address of the FCF.
FCF Interface	Slot/port number of the interface to which the FCF is connected.
VLAN	VLAN ID number used by the session.
FC-MAP	FC-Map value advertised by the FCF.
ENode Interface	Slot/number of the interface connected to the ENode.
FKA_ADV_PERIOD	Period of time (in milliseconds) during which FIP keep-alive advertisements are transmitted.
No of ENodes	Number of ENodes connected to the FCF.
FC-ID	Fibre Channel session ID assigned by the FCF.

The following example shows the `show fip-snooping statistics interface vlan` command (VLAN and port).

```
Dell# show fip-snooping statistics interface vlan 100
Number of Vlan Requests                            :0
Number of Vlan Notifications                       :0
Number of Multicast Discovery Solicits             :2
Number of Unicast Discovery Solicits               :0
Number of FLOGI                                    :2
Number of FDISC                                    :16
Number of FLOGO                                    :0
Number of Enode Keep Alive                         :9021
Number of VN Port Keep Alive                       :3349
Number of Multicast Discovery Advertisement        :4437
Number of Unicast Discovery Advertisement          :2
Number of FLOGI Accepts                            :2
Number of FLOGI Rejects                            :0
Number of FDISC Accepts                            :16
Number of FDISC Rejects                            :0
Number of FLOGO Accepts                            :0
Number of FLOGO Rejects                            :0
Number of CVL                                      :0
Number of FCF Discovery Timeouts                   :0
Number of VN Port Session Timeouts                 :0
Number of Session failures due to Hardware Config :0
Dell(conf)#

Dell# show fip-snooping statistics int tengigabitethernet 0/11
Number of Vlan Requests                            :1
Number of Vlan Notifications                       :0
Number of Multicast Discovery Solicits             :1
Number of Unicast Discovery Solicits               :0
Number of FLOGI                                   :1
Number of FDISC                                   :16
Number of FLOGO                                   :0
Number of Enode Keep Alive                        :4416
Number of VN Port Keep Alive                      :3136
Number of Multicast Discovery Advertisement       :0
Number of Unicast Discovery Advertisement         :0
Number of FLOGI Accepts                           :0
Number of FLOGI Rejects                           :0
Number of FDISC Accepts                           :0
Number of FDISC Rejects                           :0
Number of FLOGO Accepts                           :0
Number of FLOGO Rejects                           :0
Number of CVL                                     :0
Number of FCF Discovery Timeouts                  :0
Number of VN Port Session Timeouts                :0
Number of Session failures due to Hardware Config :0
```

The following example shows the `show fip-snooping statistics port-channel` command.

```
Dell# show fip-snooping statistics interface port-channel 22
Number of Vlan Requests                           :0
Number of Vlan Notifications                      :2
Number of Multicast Discovery Solicits            :0
Number of Unicast Discovery Solicits              :0
Number of FLOGI                                   :0
Number of FDISC                                   :0
Number of FLOGO                                   :0
Number of Enode Keep Alive                        :0
Number of VN Port Keep Alive                      :0
Number of Multicast Discovery Advertisement       :4451
Number of Unicast Discovery Advertisement         :2
Number of FLOGI Accepts                           :2
Number of FLOGI Rejects                           :0
Number of FDISC Accepts                           :16
Number of FDISC Rejects                           :0
Number of FLOGO Accepts                           :0
Number of FLOGO Rejects                           :0
Number of CVL                                     :0
Number of FCF Discovery Timeouts                  :0
Number of VN Port Session Timeouts                :0
Number of Session failures due to Hardware Config :0
```

The following table describes the `show fip-snooping statistics` command fields.

Table 26. `show fip-snooping statistics` Command Descriptions

Field	Description
Number of VLAN Requests	Number of FIP-snooped VLAN request frames received on the interface.
Number of VLAN Notifications	Number of FIP-snooped VLAN notification frames received on the interface.
Number of Multicast Discovery Solicits	Number of FIP-snooped multicast discovery solicit frames received on the interface.
Number of Unicast Discovery Solicits	Number of FIP-snooped unicast discovery solicit frames received on the interface.

Field	Description
Number of FLOGI	Number of FIP-snooped FLOGI request frames received on the interface.
Number of FDISC	Number of FIP-snooped FDISC request frames received on the interface.
Number of FLOGO	Number of FIP-snooped FLOGO frames received on the interface.
Number of ENode Keep Alives	Number of FIP-snooped ENode keep-alive frames received on the interface.
Number of VN Port Keep Alives	Number of FIP-snooped VN port keep-alive frames received on the interface.
Number of Multicast Discovery Advertisements	Number of FIP-snooped multicast discovery advertisements received on the interface.
Number of Unicast Discovery Advertisements	Number of FIP-snooped unicast discovery advertisements received on the interface.
Number of FLOGI Accepts	Number of FIP FLOGI accept frames received on the interface.
Number of FLOGI Rejects	Number of FIP FLOGI reject frames received on the interface.
Number of FDISC Accepts	Number of FIP FDISC accept frames received on the interface.
Number of FDISC Rejects	Number of FIP FDISC reject frames received on the interface.
Number of FLOGO Accepts	Number of FIP FLOGO accept frames received on the interface.
Number of FLOGO Rejects	Number of FIP FLOGO reject frames received on the interface.
Number of CVLs	Number of FIP clear virtual link frames received on the interface.
Number of FCF Discovery Timeouts	Number of FCF discovery timeouts that occurred on the interface.
Number of VN Port Session Timeouts	Number of VN port session timeouts that occurred on the interface.
Number of Session failures due to Hardware Config	Number of session failures due to hardware configuration that occurred on the interface.

The following example shows the `show fip-snooping system` command.

```
Dell# show fip-snooping system
Global Mode                  : Enabled
FCOE VLAN List (Operational) : 1, 100
FCFs                         : 1
Enodes                       : 2
Sessions                     : 17
```

The following example shows the `show fip-snooping vlan` command.

```
Dell# show fip-snooping vlan
* = Default VLAN

VLAN   FC-MAP     FCFs   Enodes   Sessions
----   ------     ----   ------   --------
*1     -          -      -        -
100    0X0EFC00   1      2        17
```

FCoE Transit Configuration Example

The following illustration shows an S4810 switch used as a FIP snooping bridge for FCoE traffic between an ENode (server blade) and an FCF (ToR switch). The ToR switch operates as an FCF and FCoE gateway.

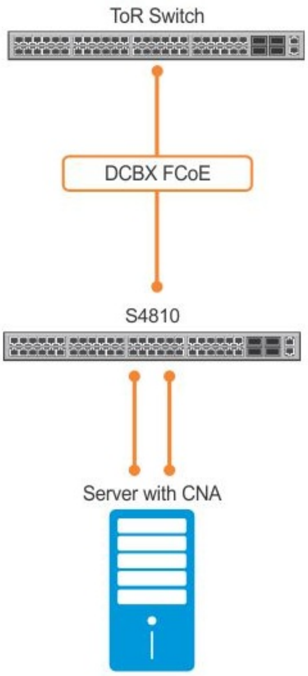


Figure 35. Configuration Example: FIP Snooping on an S4810 Switch

In this example, DCBx and PFC are enabled on the FIP snooping bridge and on the FCF ToR switch. On the FIP snooping bridge, DCBx is configured as follows:

- A server-facing port is configured for DCBx in an auto-downstream role.
- An FCF-facing port is configured for DCBx in an auto-upstream or configuration-source role.

The DCBx configuration on the FCF-facing port is detected by the server-facing port and the DCB PFC configuration on both ports is synchronized. For more information about how to configure DCBx and PFC on a port, refer to the Data Center Bridging (DCB) chapter.

The following example shows how to configure FIP snooping on FCoE VLAN 10, on an FCF-facing port (0/50), on an ENode server-facing port (0/1), and to configure the FIP snooping ports as tagged members of the FCoE VLAN enabled for FIP snooping.

Example of Enabling the FIP Snooping Feature on the Switch (FIP Snooping Bridge)

```
Dell(conf)# feature fip-snooping
```

Example of Enabling FIP Snooping on the FCoE VLAN

```
Dell(conf)# interface vlan 10
Dell(conf-if-vl-10)# fip-snooping enable
```

Example of Enabling an FC-MAP Value on a VLAN

```
Dell(conf-if-vl-10)# fip-snooping fc-map 0xOEFC01
```

NOTE: Configuring an FC-MAP value is only required if you do not use the default FC-MAP value (0x0EFC00).

Example of Configuring the ENode Server-Facing Port

```
Dell(conf)# interface tengigabitethernet 0/1
Dell(conf-if-te-0/1)# portmode hybrid
Dell(conf-if-te-0/1)# switchport
Dell(conf-if-te-0/1)# protocol lldp
Dell(conf-if-te-0/1-lldp)# dcbx port-role auto-downstream
```

NOTE: A port is enabled by default for bridge-ENode links.

Example of Configuring the FCF-Facing Port

```
Dell(conf)# interface tengigabitethernet 0/50
Dell(conf-if-te-0/50)# portmode hybrid
Dell(conf-if-te-0/50)# switchport
Dell(conf-if-te-0/50)# fip-snooping port-mode fcf
Dell(conf-if-te-0/50)# protocol lldp
Dell(conf-if-te-0/50-lldp)# dcbx port-role auto-upstream
```

Example of Configuring FIP Snooping Ports as Tagged Members of the FCoE VLAN

```
Dell(conf)# interface vlan 10
Dell(conf-if-vl-10)# tagged tengigabitethernet 0/1
Dell(conf-if-vl-10)# tagged tengigabitethernet 0/50
Dell(conf-if-te-0/1)# no shut
Dell(conf-if-te-0/50)# no shut
Dell(conf-if-vl-10)# no shut
```

After FIP packets are exchanged between the ENode and the switch, a FIP snooping session is established. ACLs are dynamically generated for FIP snooping on the FIP snooping bridge/switch.

15

Flex Hash and Optimized Boot-Up

This chapter describes the Flex Hash and fast-boot enhancements.

Flex Hash Capability Overview

This functionality is supported on the S6000 platform.

The flex hash functionality enables you to configure a packet search key and matches packets based on the search key. When a packet matches the search key, two 16-bit hash fields are extracted from the start of the L4 header and provided as inputs (bins 2 and 3) for RTAG7 hash computation. You must specify the offset of hash fields from the start of the L4 header, which contains a flow identification field.

You can configure the system to include the fields present at the offsets that you define (from the start of the L4 header) as a part of LAG and ECMP computation. Also, you can specify whether the IPv4 or IPv6 packets must be operated with the Flex Hash mechanism.

Keep the following points in mind when you configure the flex hash capability:

- A maximum of eight flex hash entries is supported.
- A maximum of 4 bytes can be extracted from the start of the L4 header.
- The offset range is 0 – 30 bytes from the start of the L4 header.
- Flex hash uses the RTAG7 bins 2 and 3 (overlay bins). These bins must be enabled for flex hash to be configured.
- If you configure flex hash by using the `load-balance ingress-port enable` and the `load-balance flexhash` commands, the `show ip flow` and `show port-channel-flow` commands are not operational. Flex hash settings and these show commands are mutually exclusive; only one of these capabilities can be functional at a time.

Configuring the Flex Hash Mechanism

This feature is supported on the S6000 platform.

The flex hash functionality enables you to configure a packet search key and matches packets based on the search key. When a packet matches the search key, two 16-bit hash fields are extracted from the start of the L4 header and provided as inputs (bins 2 and 3) for RTAG7 hash computation. You must specify the offset of hash fields from the start of the L4 header, which contains a flow identification field.

1. In Dell Networking OS Release 9.3(0.0), you can enable bins 2 and 3 by using the `load-balance ingress-port enable` command in Global Configuration mode. To configure the flex hash functionality, you must enable these bins.

   CONFIGURATION mode

   ```
   S6000-109-FTOS(conf)# load-balance ingress-port enable
   ```

When load balancing RRoCE packets using flex hash is enabled, the `show ip flow` command is disabled. Similarly, when the `show ip flow` command is in use (ingress port-based load balancing is disabled), the hashing of RRoCE packets is disabled.

Flex hash APIs do not mask out unwanted byte values after extraction of the data from the Layer 4 headers for the offset value.

2. Use the `load-balance flexhash` command to specify whether IPv4 or IPv6 packets must be subjected to the flex hash functionality, a unique protocol number, the offset of hash fields from the start of the L4 header to be used for hash calculation, and a meaningful description to associate the protocol number with the name.

   CONFIGURATION mode

   ```
   Dell(conf)# load-balance flexhash ipv4/ipv6 ip-proto <protocol number>
   <description string> offset1 <offset1 value> [offset2 <offset2 value>]
   ```

   To delete the configured flex hash setting, use the no version of the command.

Configuring Fast Boot and LACP Fast Switchover

Configure the optimized booting time functionality by performing the following steps. This procedure is supported on the S6000 platform.

1. Enable the system to restart with optimized booting-time functionality enabled.

   CONFIGURATION mode

   ```
   Dell(conf)#reload-type fastboot
   ```

2. Configure fast boot on a port-channel on both the nodes that are members of a port-channel in order to enable the physical ports to be aggregated faster. You can configure the optimal switchover functionality for LACP even if you do not enable the fast boot mode on the system. This command applies to dynamic port-channel interfaces only. When applied on a static port-channel, this command has no effect. If you configure the optimized booting-time capability and perform a reload of the system, the LACP application sends PDUs across all the active LACP links immediately.

   INTERFACE (conf-if-po-number) mode

   ```
   Dell(conf-if-po-number)#lacp fast-switchover
   ```

Optimizing the Boot Time

This functionality is supported on the S6000 platform.

You can reduce the booting time of an S6000 switch by using the fast boot feature. With the reduced time that is taken to reboot the switch, upon a manually-initiated reload or an expected restart of the device, there is minimal disruption in the traffic that is managed by the switch. Traffic outage is lowered considerably (reduced to approximately 25 seconds in certain network deployments) when you enable this optimization method while booting the device. By reducing the duration of traffic loss, the subscriber sessions are processed and preserved in an effective and seamless way.

You can configure this capability on an S6000 switch that is deployed as a top-of-rack (ToR) switch. The ToR switch is the single point of connection to the network for servers in that rack. This functionality of minimized reload time is supported in a network deployment in which the servers are connected through a ToR, leaf and spine unit or configuration setup. An exterior border gateway protocol (EBGP) session exists between the ToR and leaf switch units, and between the leaf and spine units or nodes.

Booting Process When Optimized Boot Time Mechanism is Enabled

When an S6000 switch running Dell Networking OS earlier than Release 9.3(0.0) is reloaded, the CPU and other components on the board are reset at the same time. Therefore, the control plane and the forwarding plane are impacted immediately. After the system boots up and re-initializes, the interfaces come up, control plane protocols are reestablished, and the network topology information (such as routes, adjacency settings) is learned and installed before the traffic resumes. In a typical network scenario, a traffic disconnection of 150 seconds or more usually occurs. When you employ the optimized booting functionality, the traffic outage duration is reduced drastically.

Guidelines for Configuring Optimized Booting Mechanism

Keep the following points and limitations in mind when you configure the fast boot capability:

- Fast boot is supported only when you perform an expected, stipulated reload by using the `reload-type normal-reload` command in Global Configuration mode or by using the `reset` command in uBoot mode on a switch that is running Dell Networking OS Release 9.3(0.0) or later, or when you perform a planned upgrade (and not an abrupt or unexpected shutdown) from an older release of Dell Networking OS to Release 9.3(0.0) or later. Dell recommends that you do not downgrade your system from Release 9.3(0.0) to an earlier release that does not support the fast boot functionality because the system behavior is unexpected and undefined.
- Fast boot uses the Symmetric Multiprocessing (SMP) utility that is enabled on the Intel CPU on the S6000 Switch to enhance the speed of the system startup. SMP is supported on the S6000 platform.

For the fast boot feature to reduce the traffic disruption significantly, the following conditions apply:

1. When LACP is used between the ToR switch and the adjacent devices, LACP is configured on these adjacent devices with a timeout value of 90 seconds or longer.
2. BGP timers between the ToR switch and adjacent devices are set to high values (for example, a hold timeout of 180 seconds) unless BGP graceful restart is used.
3. Before performing the planned reload, we recommend that the IPv6 Neighbor Discovery (ND) reachable timer is increased to a value of 300 seconds or longer on the adjacent devices to prevent the ND cache entries from becoming stale and being removed while the ToR goes through a CPU reset. This timer can be restored to its prior value after the ToR has completed its planned reload.
4. The BGP protocol on the adjacent devices responds to network (link-state) changes and route advertisements quickly and propagates these further up the network quickly. You might need to adjust the BGP timers on these devices.
5. Note that fast boot will operate even if some of the preceding conditions are not met. However, the duration of traffic loss might be longer.
6. Warm boot is supported because it enables faster convergence and reduced traffic loss.
7. BGP graceful restart must be configured with GR time left to default (120 seconds) or higher. The BGP hold timer should be configured as 10 seconds.
8. You must configure the LACP long timeout, which is the amount of time that a LAG interface waits for a PDU from the remote system before bringing the LACP session down, to be higher than the default value.

9. Traffic from North-South and South-North nodes are of line rate type.
10. Traffic outage for a planned reboot is less than 30 seconds for 4000 routes of IPv4 and IPv6 traffic for all of the following traffic directions. These traffic patterns apply only to the S6000 platforms.

- South-North
- North-South
- East-West
- West-East

To the south of ToR switch, 96 servers can be linked. Up to 8 Multiprocotol BGP (MP-BGP) sessions to the servers are established. You can configure a minimum of 2 MP-BGP sessions and a maximum of 8 MP-BGP sessions.

To the north of the ToR switch, up to 8 leaf nodes are connected. Up to 8 EBGP sessions for IPv4 and IPv6 for each leaf node are configured. LACP is enabled between the ToR and leaf nodes, and the LACP long timer is set to the default value. You must configure 96 ports to be 10-Gigabit Ethernet interfaces and 8 ports as 40-Gigabit Ethernet interfaces. You must configure the switch to operate with an uplink speed of 40 Gigabit Ethernet per second.

Interoperation of Applications with Fast Boot and System States

This functionality is supported on the S6000 platform.

The following sections describe the application behavior when fast boot functionality is enabled:

LACP and IPv4 Routing

Prior to the system restart, the system implements the following changes when you perform a fast boot:

The system saves all dynamic ARP entries to a database on the flash drive.

A file is generated to indicate that the system is undergoing a fast boot, which is used after the system comes up.

After the Dell Networking OS image is loaded and activated, and the appropriate software components come up, the following additional actions are performed:

- If a database of dynamic ARP entries is present on the flash drive, that information is read and the ARP entries are restored; the entries are installed on the switch as soon as possible. At the same time, the entries are changed to an initial ("aged out") state so that they are refreshed (and flushed if not learnt again). The database on the flash card is also deleted instantaneously.
- The system ensures that local routes known to BGP are imported into BGP and advertised to peers as quickly as possible. In this process, any advertisement-interval configuration is not considered (only during the initial period when the peer comes up).

If you do not configure BGP GR, you must configure the peering with BGP keepalive and hold timers to be as high as possible (depending on your network deployment and the scaled parameters or sessions) to enable the connection to be active until the system re-initializes the switch, causing the links to adjacent devices to go down. If the BGP sessions are disabled before the re-initialization of the switch occurs because of the peer timing out, traffic disruption occurs from that point onwards, even if the system continues to maintain valid routing information in the hardware and is capable of forwarding traffic.

LACP and IPv6 Routing

The following IPv6-related actions are performed during the reload phase:

- The system saves all the dynamic ND cache entries to a database on the flash card. After the system comes back online, and the Dell Networking OS image is loaded and the corresponding software applications on the system are also activated, the following processes specific to IPv6 are performed:
- If a database of dynamic ND entries is present on the flash, the information is read and the ND entries are restored (to the IPv6 subsystem as well as the kernel); the entries are installed on the switch as quickly as possible. At the same time, the entries are changed to an initial ("incomplete") state so that they are refreshed (and flushed, if not learnt again). The database on the flash is also deleted immediately.
- To ensure that the adjacent systems do not time out and purge their ND cache entries, the age-out time or the reachable time for ND cache entries must be configured to be as high as necessary. Dell recommends that you configure the reachable timer to be 90 seconds or longer.

BGP Graceful Restart

When the system contains one or more BGP peerings configured for BGP graceful restart, fast boot performs the following actions:

- A closure of the TCP sessions is performed on all sockets corresponding to BGP sessions on which Graceful Restart has been negotiated. This behavior is to force the peer to perform the helper role so that any routes advertised by the restarting system are retained and the peering session will not go down due to BGP Hold timeout.
- Termination of TCP connections is not initiated on BGP sessions without GR because such a closure might cause the peer to immediately purge routes learnt from the restarting system.
- When BGP is started, it sets the R-bit and F-bit in the GR capability when bringing up the session with peers for which BGP GR has been configured. This is the standard behavior of a restarting system and ensures that the peer continues to retain the routes previously advertised by the system.
- The system delays sending the BGP End-of-RIB notification to peers with whom BGP GR has been negotiated to ensure that the local routes of the system are advertised to the peers, if required by the configuration.
- If BGP GR is enabled on any peering session, the timeout values used for the BGP hold timer do not take effect.

Cold Boot Caused by Power Cycling the System

When you perform a power-cycle operation on a system that is configured with the optimized booting functionality, the system goes through its regular boot sequence even if it is configured for fast boot. When the system comes up, it is expected that there will be no dynamic ARP or ND database to restore. The system boot up mode will not be fast boot and

Unexpected Reload of the System

When an unexpected or unplanned reload occurs, such as a reset caused by the software, the system performs the regular boot sequence even if it is configured for fast boot. When the system comes up, dynamic ARP or ND database entries are not present or required to be restored. The system boot up mode will not be fast boot and actions specific to this mode will not be performed.

Software Upgrade

When fast boot is used to upgrade the system to a release that supports fast boot, the system enables the restoration of dynamic ARP or ND databases that were maintained in the older release from when you performed the upgrade and the ARP and ND applications identify that the system has been booted using fast boot.

LACP Fast Switchover

For fast boot, the operation of LACP has been optimized. These LACP optimizations are applicable even when fast boot is not enabled when a system reload is performed. These enhancements are controlled using the fast-switchover option that is available with the `lacp` command in Port Channel Interface Configuration mode. When LACP 'fast-switchover' is enabled on the system, two optimizations are performed to the LACP behavior:

- The wait-while timer is not started in the 'waiting' state of the MUX state machine. The port moves directly to the 'attached' state.
- The local system moves to the 'collecting' and 'distributing' states on the port in a single step without waiting for the partner to set the 'collecting' bit.

Changes to BGP Multipath

When the system becomes active after a fast-boot restart, a change has been made to the BGP multipath and ECMP behavior. The system delays the computation and installation of additional paths to a destination into the BGP routing information base (RIB) and forwarding table for a certain period of time. Additional paths, if any, are automatically computed and installed without the need for any manual intervention in any of the following conditions:

- After 30 seconds of the system returning online after a restart
- After all established peers have synchronized with the restarting system
- A combination of the previous two conditions

One possible impact of this behavior change is that if the amount of traffic to a destination is higher than the volume of traffic that can be carried over one path, a portion of that traffic might be dropped for a short duration (30-60 seconds) after the system comes up.

Delayed Installation of ECMP Routes Into BGP

The current FIB component of Dell Networking OS has some inherent inefficiencies when handling a large number of ECMP routes (i.e., routes with multiple equal-cost next hops). To circumvent this for the configuration of fast boot, changes are made in BGP to delay the installation of ECMP routes. This is done only if the system comes up through a fast boot reload. The BGP route selection algorithm only selects one best path to each destination and delays installation of additional ECMP paths until a minimum of 30 seconds has elapsed from the time the first BGP peer is established. Once this time has elapsed, all routes in the BGP RIB are processed for additional paths.

While the above change will ensure that at least one path to each destination gets into the FIB as quickly as possible, it does prevent additional paths from being used even if they are available. This downside has been deemed to be acceptable.

RDMA Over Converged Ethernet (RoCE) Overview

This functionality is supported on the S6000 platform.

Remote direct memory access (RDMA) reduces both CPU cycles and latency. RDMA over converged Ethernet (RoCE) implements IB over Ethernet. RRoCE sends InfiniBand (IB) packets over IP. IB supports input and output connectivity for the internet infrastructure. InfiniBand enables the expansion of network topologies over large geographical boundaries and the creation of next-generation I/O interconnect standards in servers. Although the endpoints or the destination servers generate such RRoCE packets, from the switch's perspective, RRoCE is processed as an IP packet.

RRoCE packets are received and transmitted on specific interfaces called lite-subinterfaces. These interfaces are similar to the normal Layer 3 physical interfaces except for the extra provisioning that they offer to enable the VLAN ID for encapsulation.

You can configure a physical interface or a Layer 3 Port Channel interface as a lite subinterface. When you configure a lite subinterface, only tagged IP packets with VLAN encapsulation are processed and routed. All other data packets are discarded.

To provide lossless service for RRoCE, the QoS service policy must be configured in the ingress and egress directions on lite subinterfaces.

A normal Layer 3 physical interface processes only untagged packets and makes routing decisions based on the default Layer 3 VLAN ID (4095).

To enable routing of RRoCE packets, the VLAN ID is mapped to the default VLAN ID of 4095 using VLAN translation. After the VLAN translation, the RRoCE packets are processed in the same way as normal IP packets that a Layer 3 interface receives and routes in the egress direction. At the egress interface, the VLAN ID is appended to the packet and transmitted out of the interface as a tagged packet with the dot1Q value preserved.

RDMA is a technology that a virtual machine (VM) uses to directly transfer information to the memory of another VM, thus enabling VMs to be connected to storage networks. With RoCE, RDMA enables data to be forwarded without passing through the CPU and the main memory path of TCP/IP. In a deployment that contains both the RoCE network and the normal IP network on two different networks, RRoCE combines the RoCE and the IP networks and sends the RoCE frames over the IP network. This method of transmission, called RRoCE, results in the encapsulation of RoCE packets to IP packets.

When a storage area network (SAN) is connected over an IP network, the following conditions must be satisfied:

- Faster Connectivity: QoS for RRoCE enables faster and lossless nature of disk input and output services.
- Lossless connectivity: VMs require the connectivity to the storage network to be lossless always. When a planned upgrade of the network nodes happens, especially with top-of-rack (ToR) nodes where there is a single point of failure for the VMs, disk I/O operations are expected to occur in 20 seconds. If disk in not accessible in 20 seconds, unexpected and undefined behavior of the VMs

occurs. You can optimize the booting time of the ToR nodes that experience a single point of failure to reduce the outage in traffic-handling operations.

RoCE over a routed system is called RRoCE. RRoCE has IP headers. RRoCE is bursty and uses the entire 10-Gigabit Ethernet interface. Although RRoCE and normal data traffic are propagated in separate network portions, it may be necessary in certain topologies to combine both the RRoCE and the data traffic in a single network structure. RRoCE traffic is marked with dot1p priorities 3 and 4 (code points 011 and 100, respectively) and these queues are strict and lossless. DSCP code points are not tagged for RRoCE. Both ECN and PFC are enabled for RRoCE traffic. For normal IP or data traffic that is not RRoCE-enabled, the packets comprise TCP and UDP packets and they can be marked with DSCP code points. Multicast is not supported in that network.

Preserving 802.1Q VLAN Tag Value for Lite Subinterfaces

This functionality is supported on the S6000 platform.

All the frames in a Layer 2 VLAN are identified using a tag defined in the IEEE 802.1Q standard to determine the VLAN to which the frames or traffic are relevant or associated. Such frames are encapsulated with the 802.1Q tags. If a single VLAN is configured in a network topology, all the traffic packets contain the same do1q tag, which is the tag value of the 802.1Q header. If a VLAN is split into multiple, different sub-VLANs, each VLAN is denoted by a unique 8021.Q tag to enable the nodes that receive the traffic frames determine the VLAN for which the frames are destined.

Typically, a Layer 3 physical interface processes only untagged or priority-tagged packets. Tagged packets that are received on Layer 3 physical interfaces are dropped. To enable the routing of tagged packets, the port that receives such tagged packets needs to be configured as a switchport and must be bound to a VLAN as a tagged member port.

A lite subinterface is similar to a normal Layer 3 physical interface, except that additional provisioning is performed to set the VLAN ID for encapsulation.

A physical interface or a Layer 3 Port channel interface can be configured as a lite subinterface. Once a lite subinterface is configured, only tagged IP packets with encapsulation VLAN ID are processed and routed. All other data packets are discarded except the Layer 2 and Layer 3 control frames. It is not required for a VLAN ID to be preserved (in the hardware or the OS application) when a VLAN ID, used for encapsulation, is associated with a physical/Port-channel interface. Normal VLANs and VLAN encapsulation can exist simultaneously and any non-unicast traffic received on a normal VLAN is not flooded using lite subinterfaces whose encapsulation VLAN ID matches with that of the normal VLAN ID.

You can use the `encapsulation dot1q` *`vlan-id`* command in INTERFACE mode to configure lite subinterfaces.

16

Force10 Resilient Ring Protocol (FRRP)

Force10 resilient ring protocol (FRRP) is supported on the S6000 platform.

FRRP provides fast network convergence to Layer 2 switches interconnected in a ring topology, such as a metropolitan area network (MAN) or large campuses. FRRP is similar to what can be achieved with the spanning tree protocol (STP), though even with optimizations, STP can take up to 50 seconds to converge (depending on the size of network and node of failure) may require 4 to 5 seconds to reconverge. FRRP can converge within 150ms to 1500ms when a link in the ring breaks (depending on network configuration).

To operate a deterministic network, a network administrator must run a protocol that converges independently of the network size or node of failure. FRRP is a proprietary protocol that provides this flexibility, while preventing Layer 2 loops. FRRP provides sub-second ring-failure detection and convergence/re-convergence in a Layer 2 network while eliminating the need for running spanning-tree protocol. With its two-way path to destination configuration, FRRP provides protection against any single link/switch failure and thus provides for greater network uptime.

Protocol Overview

FRRP is built on a ring topology.

You can configure up to 255 rings on a system. FRRP uses one Master node and multiple Transit nodes in each ring. There is no limit to the number of nodes on a ring. The Master node is responsible for the intelligence of the Ring and monitors the status of the Ring. The Master node checks the status of the Ring by sending ring health frames (RHF) around the Ring from its Primary port and returning on its Secondary port. If the Master node misses three consecutive RHFs, the Master node determines the ring to be in a failed state. The Master then sends a Topology Change RHF to the Transit Nodes informing them that the ring has changed. This causes the Transit Nodes to flush their forwarding tables, and re-converge to the new network structure.

One port of the Master node is designated the Primary port (P) to the ring; another port is designated as the Secondary port (S) to the ring. In normal operation, the Master node blocks the Secondary port for all non-control traffic belonging to this FRRP group, thereby avoiding a loop in the ring, like STP. Layer 2 switching and learning mechanisms operate per existing standards on this ring.

Each Transit node is also configured with a Primary port and a Secondary port on the ring, but the port distinction is ignored as long as the node is configured as a Transit node. If the ring is complete, the Master node logically blocks all data traffic in the transmit and receive directions on the Secondary port to prevent a loop. If the Master node detects a break in the ring, it unblocks its Secondary port and allows data traffic to be transmitted and received through it. Refer to the following illustration for a simple example of this FRRP topology. Note that ring direction is determined by the Master node’s Primary and Secondary ports.

A virtual LAN (VLAN) is configured on all node ports in the ring. All ring ports must be members of the Member VLAN and the Control VLAN.

The Member VLAN is the VLAN used to transmit data as described earlier.

The Control VLAN is used to perform the health checks on the ring. The Control VLAN can always pass through all ports in the ring, including the secondary port of the Master node.

Ring Status

The ring failure notification and the ring status checks provide two ways to ensure the ring remains up and active in the event of a switch or port failure.

Ring Checking

At specified intervals, the Master node sends a ring health frame (RHF) through the ring. If the ring is complete, the frame is received on its secondary port and the Master node resets its fail-period timer and continues normal operation.

If the Master node does not receive the RHF before the fail-period timer expires (a configurable timer), the Master node moves from the Normal state to the Ring-Fault state and unblocks its Secondary port. The Master node also clears its forwarding table and sends a control frame to all other nodes, instructing them to also clear their forwarding tables. Immediately after clearing its forwarding table, each node starts learning the new topology.

Ring Failure

If a Transit node detects a link down on any of its ports on the FRRP ring, it immediately sends a link-down control frame on the Control VLAN to the Master node.

When the Master node receives this control frame, the Master node moves from the Normal state to the Ring-Fault state and unblocks its Secondary port. The Master node clears its routing table and sends a control frame to all other ring nodes, instructing them to clear their routing tables as well. Immediately after clearing its routing table, each node begins learning the new topology.

Ring Restoration

The Master node continues sending ring health frames out its primary port even when operating in the Ring-Fault state.

After the ring is restored, the next status check frame is received on the Master node's Secondary port. This causes the Master node to transition back to the Normal state. The Master node then logically blocks non-control frames on the Secondary port, clears its own forwarding table, and sends a control frame to the Transit nodes, instructing them to clear their forwarding tables and re-learn the topology.

During the time between the Transit node detecting that its link is restored and the Master node detecting that the ring is restored, the Master node’s Secondary port is still forwarding traffic. This can create a temporary loop in the topology. To prevent this, the Transit node places all the ring ports transiting the newly restored port into a temporary blocked state. The Transit node remembers which port has been temporarily blocked and places it into a pre- forwarding state. When the Transit node in the pre-forwarding state receives the control frame instructing it to clear its routing table, it does so and unblocks the previously blocked ring ports on the newly restored port. Then the Transit node returns to the Normal state.

Multiple FRRP Rings

Up to 255 rings are allowed per system and multiple rings can be run on one system.

More than the recommended number of rings may cause interface instability. You can configure multiple rings with a single switch connection; a single ring can have multiple FRRP groups; multiple rings can be connected with a common link.

Member VLAN Spanning Two Rings Connected by One Switch

A member VLAN can span two rings interconnected by a common switch, in a figure-eight style topology.

A switch can act as a Master node for one FRRP group and a Transit for another FRRP group, or it can be a Transit node for both rings.

In the following example, FRRP 101 is a ring with its own Control VLAN, and FRRP 202 has its own Control VLAN running on another ring. A Member VLAN that spans both rings is added as a Member VLAN to both FRRP groups. Switch R3 has two instances of FRRP running on it: one for each ring. The example topology that follows shows R3 assuming the role of a Transit node for both FRRP 101 and FRRP 202.

Important FRRP Points

FRRP provides a convergence time that can generally range between 150ms and 1500ms for Layer 2 networks.

The Master node originates a high-speed frame that circulates around the ring. This frame, appropriately, sets up or breaks down the ring.

- The Master node transmits ring status check frames at specified intervals.
- You can run multiple physical rings on the same switch.
- One Master node per ring — all other nodes are Transit.
- Each node has two member interfaces — primary and secondary.
- There is no limit to the number of nodes on a ring.
- Master node ring port states — blocking, pre-forwarding, forwarding, and disabled.
- Transit node ring port states — blocking, pre-forwarding, forwarding, and disabled.
- STP disabled on ring interfaces.
- Master node secondary port is in blocking state during Normal operation.
- Ring health frames (RHF)
  - Hello RHF: sent at 500ms (hello interval); Only the Master node transmits and processes these.
  - Topology Change RHF: triggered updates; processed at all nodes.

Important FRRP Concepts

The following table lists some important FRRP concepts.

Concept	Explanation
Ring ID	Each *ring* has a unique 8-bit ring ID through which the ring is identified (for example, FRRP 101 and FRRP 202, as shown in the illustration in Member VLAN Spanning Two Rings Connected by One Switch.

Concept	Explanation
Control VLAN	Each *ring* has a unique Control VLAN through which tagged ring health frames (RHF) are sent. Control VLANs are used only for sending RHF, and cannot be used for any other purpose.
Member VLAN	Each *ring* maintains a list of member VLANs. Member VLANs must be consistent across the entire ring.
Port Role	Each *node* has two ports for each ring: Primary and Secondary. The Master node Primary port generates RHFs. The Master node Secondary port receives the RHFs. On Transit nodes, there is no distinction between a Primary and Secondary interface when operating in the Normal state.
Ring Interface State	Each interface (port) that is part of the ring maintains one of four states” • **Blocking State** — Accepts ring protocol packets but blocks data packets. LLDP, FEFD, or other Layer 2 control packets are accepted. Only the Master node Secondary port can enter this state. • **Pre-Forwarding State** — A transition state before moving to the Forward state. Control traffic is forwarded but data traffic is blocked. The Master node Secondary port transitions through this state during ring bring-up. All ports transition through this state when a port comes up. • **Pre-Forwarding State** — A transition state before moving to the Forward state. Control traffic is forwarded but data traffic is blocked. The Master node Secondary port transitions through this state during ring bring-up. All ports transition through this state when a port comes up. • **Disabled State** — When the port is disabled or down, or is not on the VLAN.
Ring Protocol Timers	• **Hello Interval** — The interval when ring frames are generated from the Master node’s Primary interface (default **500 ms**). The Hello interval is configurable in 50 ms increments from 50 ms to 2000 ms. • **Dead Interval** — The interval when data traffic is blocked on a port. The default is three times the Hello interval rate. The dead interval is configurable in 50 ms increments from 50 ms to 6000 ms.
Ring Status	The state of the FRRP ring. During initialization/configuration, the default ring status is Ring-down (disabled). The Primary and Secondary interfaces, control VLAN, and Master and Transit node information must be configured for the ring to be up. • **Ring-Up** — Ring is up and operational. • **Ring-Down** — Ring is broken or not set up.
Ring Health-Check Frame (RHF)	The Master node generates two types of RHFs. RHFs never loop the ring because they terminate at the Master node’s secondary port. • **Hello RHF (HRHF)** — These frames are processed only on the Master node’s Secondary port. The Transit nodes pass the HRHF through without processing it. An HRHF is sent at every Hello interval. • **Topology Change RHF (TCRHF)** — These frames contains ring status, keepalive, and the control and member VLAN hash. The TCRHF is processed at each node of the ring. TCRHFs are sent out the Master Node’s Primary and Secondary interface when the ring is declared in a Failed state with the same sequence number, on any topology change to ensure that all Transit nodes receive it.

Concept	Explanation
	There is no periodic transmission of TCRHFs. The TCRHFs are sent on triggered events of ring failure or ring restoration only.

Implementing FRRP

- FRRP is media and speed independent.
- FRRP is a Dell proprietary protocol that does not interoperate with any other vendor.
- You must disable the spanning tree protocol (STP) on both the Primary and Secondary interfaces before you can enable FRRP.
- All ring ports must be Layer 2 ports. This is required for both Master and Transit nodes.
- A VLAN configured as a control VLAN for a ring cannot be configured as a control or member VLAN for any other ring.
- The control VLAN is not used to carry any data traffic; it carries only RHFs.
- The control VLAN cannot have members that are not ring ports.
- If multiple rings share one or more member VLANs, they cannot share any links between them.
- Member VLANs across multiple rings are not supported in Master nodes.
- Each ring has only one Master node; all others are transit nodes.

FRRP Configuration

These are the tasks to configure FRRP.

- Creating the FRRP Group
- Configuring the Control VLAN
  - Configure Primary and Secondary ports
- Configuring and Adding the Member VLANs
  - Configure Primary and Secondary ports

Other FRRP related commands are:

- Clearing the FRRP Counters
- Viewing the FRRP Configuration
- Viewing the FRRP Information

Creating the FRRP Group

Create the FRRP group on each switch in the ring.
To create the FRRP group, use the command.

- Create the FRRP group with this Ring ID.

  CONFIGURATION mode

  ```
  protocol frrp ring-id
  ```

  Ring ID: the range is from 1 to 255.

Configuring the Control VLAN

Control and member VLANS are configured normally for Layer 2. Their status as control or member is determined at the FRRP group commands.
For more information about configuring VLANS in Layer 2 mode, refer to Layer 2.
Be sure to follow these guidelines:

- All VLANS must be in Layer 2 mode.
- You can only add ring nodes to the VLAN.
- A control VLAN can belong to one FRRP group only.
- Tag control VLAN ports.
- All ports on the ring must use the same VLAN ID for the control VLAN.
- You cannot configure a VLAN as both a control VLAN and member VLAN on the same ring.
- Only two interfaces can be members of a control VLAN (the Master Primary and Secondary ports).
- Member VLANs across multiple rings are not supported in Master nodes.

To create the control VLAN for this FRRP group, use the following commands on the switch that is to act as the Master node.

1. Create a VLAN with this ID number.
   CONFIGURATION mode.

   ```
   interface vlan vlan-id
   ```

   VLAN ID: from 1 to 4094.
2. Tag the specified interface or range of interfaces to this VLAN.
   CONFIG-INT-VLAN mode.

   ```
   tagged interface slot/ port {range}
   ```

   `Interface`:

   - For a 10/100/1000 Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
   - For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
   - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
   - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.

   `Slot/Port, Range`: Slot and Port ID for the interface. Range is entered Slot/Port-Port.
3. Assign the Primary and Secondary ports and the control VLAN for the ports on the ring.
   CONFIG-FRRP mode.

   ```
   interface primary int slot/port secondary int slot/port control-vlan vlan id
   ```

   `Interface`:

   - For a 10/100/1000 Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.

- For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
- For a SONET interface, enter the keyword `sonet` then the slot/port information.
- For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
- For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.

`Slot/Port, Range`: Slot and Port ID for the interface. Range is entered Slot/Port-Port.

`VLAN ID`: The VLAN identification of the control VLAN.

4. Configure the Master node.

   CONFIG-FRRP mode.

   ```
   mode master
   ```
5. Identify the Member VLANs for this FRRP group.

   CONFIG-FRRP mode.

   ```
   member-vlan vlan-id {range}
   ```

   `VLAN-ID, Range`: VLAN IDs for the ring's member VLANS.
6. Enable FRRP.

   CONFIG-FRRP mode.

   ```
   no disable
   ```

Configuring and Adding the Member VLANs

Control and member VLANS are configured normally for Layer 2. Their status as Control or Member is determined at the FRRP group commands.
For more information about configuring VLANS in Layer 2 mode, refer to the Layer 2 chapter.
Be sure to follow these guidelines:

- All VLANS must be in Layer 2 mode.
- Tag control VLAN ports. Member VLAN ports, except the Primary/Secondary interface, can be tagged or untagged.
- The control VLAN must be the same for all nodes on the ring.

To create the Members VLANs for this FRRP group, use the following commands on all of the Transit switches in the ring.

1. Create a VLAN with this ID number.

   CONFIGURATION mode.

   ```
   interface vlan vlan-id
   ```

   VLAN ID: the range is from 1 to 4094.
2. Tag the specified interface or range of interfaces to this VLAN.

   CONFIG-INT-VLAN mode.

   ```
   tagged interface slot/port {range}
   ```

   `Interface`:

- *Slot/Port, range*: Slot and Port ID for the interface. The range is entered Slot/Port-Port.
- For a 10/100/1000 Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
- For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
- For a SONET interface, enter the keyword `sonet` then the slot/port information.
- For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
- For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.

3. Assign the Primary and Secondary ports and the Control VLAN for the ports on the ring.

   CONFIG-FRRP mode.

   ```
   interface primary int slot/port secondary int slot/port control-vlan vlan id
   ```

   Interface:

   - For a 10/100/1000 Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
   - For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
   - For a SONET interface, enter the keyword `sonet` then the slot/port information.
   - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
   - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.

   Slot/Port, Range: Slot and Port ID for the interface. Range is entered Slot/Port-Port.

   VLAN ID: Identification number of the Control VLAN.

4. Configure a Transit node.

   CONFIG-FRRP mode.

   ```
   mode transit
   ```

5. Identify the Member VLANs for this FRRP group.

   CONFIG-FRRP mode.

   ```
   member-vlan vlan-id {range}
   ```

   VLAN-ID, Range: VLAN IDs for the ring's Member VLANs.

6. Enable this FRRP group on this switch.

   CONFIG-FRRP mode.

   ```
   no disable
   ```

Setting the FRRP Timers

To set the FRRP timers, use the following command.

NOTE: Set the Dead-Interval time 3 times the Hello-Interval.

- Enter the desired intervals for Hello-Interval or Dead-Interval times.
  CONFIG-FRRP mode.

  ```
  timer {hello-interval|dead-interval} milliseconds
  ```

  - *`Hello-Interval`*: the range is from 50 to 2000, in increments of 50 (default is **500**).
  - *`Dead-Interval`*: the range is from 50 to 6000, in increments of 50 (default is **1500**).

Clearing the FRRP Counters

To clear the FRRP counters, use one of the following commands.

- Clear the counters associated with this Ring ID.
  EXEC PRIVELEGED mode.

  ```
  clear frrp ring-id
  ```

  Ring ID: the range is from 1 to 255.
- Clear the counters associated with all FRRP groups.
  EXEC PRIVELEGED mode.

  ```
  clear frrp
  ```

Viewing the FRRP Configuration

To view the configuration for the FRRP group, use the following command.

- Show the configuration for this FRRP group.
  CONFIG-FRRP mode.

  ```
  show configuration
  ```

Viewing the FRRP Information

To view general FRRP information, use one of the following commands.

- Show the information for the identified FRRP group.
  EXEC or EXEC PRIVELEGED mode.

  ```
  show frrp ring-id
  ```

  Ring ID: the range is from 1 to 255.
- Show the state of all FRRP groups.
  EXEC or EXEC PRIVELEGED mode.

  ```
  show frrp summary
  ```

  Ring ID: the range is from 1 to 255.

Troubleshooting FRRP

To troubleshoot FRRP, use the following information.

Configuration Checks

- Each Control Ring must use a unique VLAN ID.
- Only two interfaces on a switch can be Members of the same control VLAN.
- There can be only one Master node for any FRRP group.
- You can configure FRRP on Layer 2 interfaces only.
- Spanning Tree (if you enable it globally) must be disabled on both Primary and Secondary interfaces when you enable FRRP.
  - When the interface ceases to be a part of any FRRP process, if you enable Spanning Tree globally, also enable it explicitly for the interface.
- The maximum number of rings allowed on a chassis is 255.

Sample Configuration and Topology

The following example shows a basic FRRP topology.

Example of R1 MASTER

```
interface GigabitEthernet 1/24
  no ip address
  switchport
  no shutdown
!
interface GigabitEthernet 1/34
  no ip address
  switchport
  no shutdown
!
interface Vlan 101
  no ip address
  tagged GigabitEthernet 1/24,34
  no shutdown
!
interface Vlan 201
  no ip address
  tagged GigabitEthernet 1/24,34
  no shutdown

!
protocol frrp 101
  interface primary GigabitEthernet 1/24
secondary GigabitEthernet 1/34 control-vlan 101
  member-vlan 201
  mode master
  no disable
```

Example of R2 TRANSIT

```
interface GigabitEthernet 2/14
  no ip address
  switchport
  no shutdown
!
interface GigabitEthernet 2/31
```

```
  no ip address
  switchport
  no shutdown
!
interface Vlan 101
  no ip address
  tagged GigabitEthernet 2/14,31
  no shutdown
!
interface Vlan 201
  no ip address
  tagged GigabitEthernet 2/14,31
  no shutdown
!
protocol frrp 101
  interface primary GigabitEthernet 2/14 secondary GigabitEthernet 2/31 control-
vlan 101
  member-vlan 201
  mode transit
  no disable
```

Example of R3 TRANSIT

```
interface GigabitEthernet 3/14
  no ip address
  switchport
  no shutdown
!
interface GigabitEthernet 3/21
  no ip address
  switchport
  no shutdown
!
interface Vlan 101
  no ip address
  tagged GigabitEthernet 3/14,21
  no shutdown
!
interface Vlan 201
  no ip address
  tagged GigabitEthernet 3/14,21
  no shutdown

!
protocol frrp 101
  interface primary GigabitEthernet 3/21
secondary GigabitEthernet 3/14 control-vlan 101
  member-vlan 201
  mode transit
  no disable
```

17

GARP VLAN Registration Protocol (GVRP)

GARP VLAN registration protocol (GVRP) is supported on the S6000 platform.

Typical virtual local area network (VLAN) implementation involves manually configuring each Layer 2 switch that participates in a given VLAN. GVRP, defined by the IEEE 802.1q specification, is a Layer 2 network protocol that provides for automatic VLAN configuration of switches. GVRP-compliant switches use GARP to register and de-register attribute values, such as VLAN IDs, with each other.

GVRP exchanges network VLAN information to allow switches to dynamically forward frames for one or more VLANs. Therefore, GVRP spreads this information and configures the needed VLANs on any additional switches in the network. Data propagates via the exchange of GVRP protocol data units (PDUs).

The purpose of GVRP is to simplify (but not eliminate) static configuration. The idea is to configure switches at the edge and have the information dynamically propagate into the core. As such, the edge ports must still be statically configured with VLAN membership information, and they do not run GVRP. It is this information that is propagated to create dynamic VLAN membership in the core of the network.

Important Points to Remember

- GVRP propagates VLAN membership throughout a network. GVRP allows end stations and switches to issue and revoke declarations relating to VLAN membership.
- VLAN registration is made in the context of the port that receives the GARP PDU and is propagated to the other active ports.
- GVRP is disabled by default; enable GVRP for the switch and then for individual ports.
- Dynamic VLANs are aged out after the LeaveAll timer expires three times without receipt of a Join message. To display status, use the `show gvrp statistics {interface` *`interface`* `| summary}` command.

```
Dell(conf)#protocol spanning-tree pvst
Dell(conf-pvst)#no disable
% Error: GVRP running. Cannot enable PVST.

.........
Dell(conf)#protocol spanning-tree mstp
Dell(conf-mstp)#no disable
% Error: GVRP running. Cannot enable MSTP.

.........

Dell(conf)#protocol gvrp
Dell(conf-gvrp)#no disable
% Error: PVST running. Cannot enable GVRP.
% Error: MSTP running. Cannot enable GVRP.
```

Configure GVRP

To begin, enable GVRP.

To facilitate GVRP communications, enable GVRP globally on each switch. Then, GVRP configuration is per interface on a switch-by-switch basis. Enable GVRP on each port that connects to a switch where you want GVRP information exchanged. In the following example, that type of port is referred to as a VLAN trunk port, but it is not necessary to specifically identify to the Dell Networking OS that the port is a trunk port.

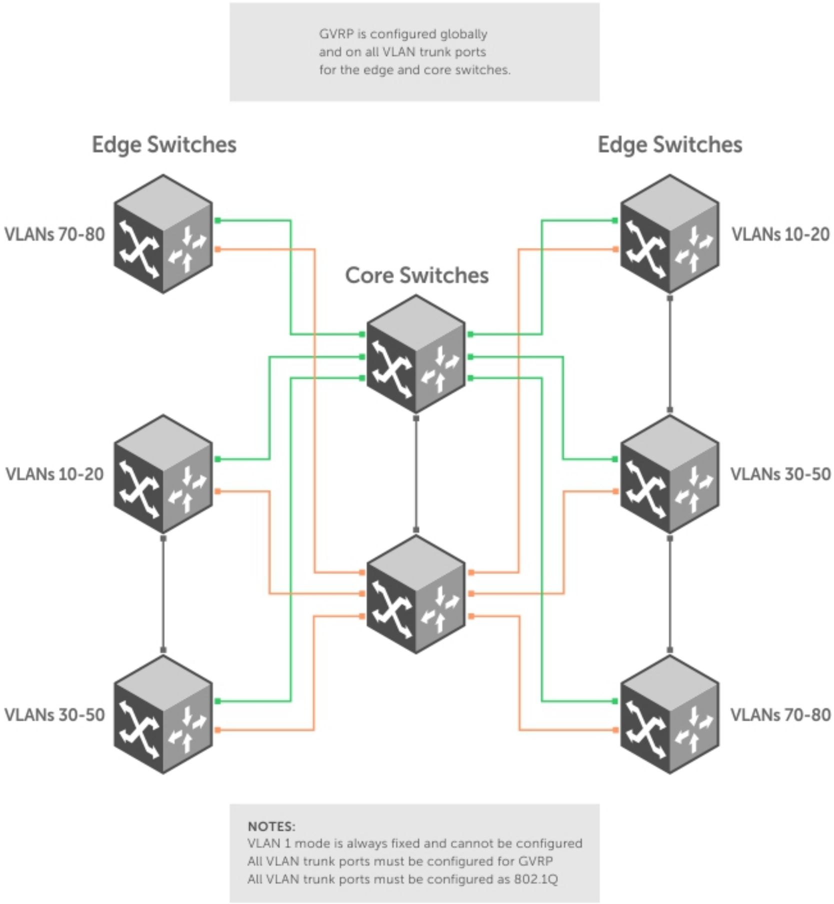

GVRP is configured globally
and on all VLAN trunk ports
for the edge and core switches.
Edge Switches
Edge Switches
VLANs 70-80
VLANs 10-20
Core Switches
VLANs 10-20
VLANs 30-50
VLANs 30-50
VLANs 70-80
NOTES:
VLAN 1 mode is always fixed and cannot be configured
All VLAN trunk ports must be configured for GVRP
All VLAN trunk ports must be configured as 802.1Q

Figure 36. Global GVRP Configuration Example

Basic GVRP configuration is a two-step process:

1. Enabling GVRP Globally
2. Enabling GVRP on a Layer 2 Interface

Related Configuration Tasks

- Configure GVRP Registration

- Configure a GARP Timer

Enabling GVRP Globally

To configure GVRP globally, use the following command.

- Enable GVRP for the entire switch.

  CONFIGURATION mode

  ```
  gvrp enable
  ```

Example of Configuring GVRP

```
Dell(conf)#protocol gvrp
Dell(config-gvrp)#no disable
Dell(config-gvrp)#show config
!
protocol gvrp
no disable
Dell(config-gvrp)#
```

To inspect the global configuration, use the `show gvrp brief` command.

Enabling GVRP on a Layer 2 Interface

To enable GVRP on a Layer 2 interface, use the following command.

- Enable GVRP on a Layer 2 interface.

  INTERFACE mode

  ```
  gvrp enable
  ```

Example of Enabling GVRP on an Interface

```
Dell(conf-if-gi-1/21)#switchport
Dell(conf-if-gi-1/21)#gvrp enable
Dell(conf-if-gi-1/21)#no shutdown
Dell(conf-if-gi-1/21)#show config
!
interface GigabitEthernet 1/21
no ip address
switchport
gvrp enable
no shutdown
```

To inspect the interface configuration, use the `show config` command from INTERFACE mode or use the `show gvrp interface` command in EXEC or EXEC Privilege mode.

Configure GVRP Registration

Configure GVRP registration.

There are two GVRP registration modes:

- **Fixed Registration Mode** — figuring a port in fixed registration mode allows for manual creation and registration of VLANs, prevents VLAN deregistration, and registers all VLANs known on other ports on the port. For example, if an interface is statically configured via the CLI to belong to a VLAN, it should not be unconfigured when it receives a Leave PDU. Therefore, the registration mode on that interface is FIXED.

- **Forbidden Mode** — Disables the port to dynamically register VLANs and to propagate VLAN information except information about VLAN 1. A port with forbidden registration type thus allows only VLAN 1 to pass through even though the PDU carries information for more VLANs. Therefore, if you do not want the interface to advertise or learn about particular VLANS, set the interface to the registration mode of FORBIDDEN.

Based on the configuration in the following example, the interface 1/21 is not removed from VLAN 34 or VLAN 35 despite receiving a GVRP Leave message. Additionally, the interface is not dynamically added to VLAN 45 or VLAN 46, even if a GVRP Join message is received.

Example of the `gvrp registration` Command

```
Dell(conf-if-gi-1/21)#gvrp registration fixed 34,35
Dell(conf-if-gi-1/21)#gvrp registration forbidden 45,46
Dell(conf-if-gi-1/21)#show conf
!
interface GigabitEthernet 1/21
  no ip address
  switchport
  gvrp enable
  gvrp registration fixed 34-35
  gvrp registration forbidden 45-46
  no shutdown
Dell(conf-if-gi-1/21)#
```

Configure a GARP Timer

Set GARP timers to the same values on all devices that are exchanging information using GVRP.

There are three GARP timer settings.

- **Join** — A GARP device reliably transmits Join messages to other devices by sending each Join message two times. To define the interval between the two sending operations of each Join message, use this parameter. The Dell Networking OS default is **200ms**.
- **Leave** — When a GARP device expects to de-register a piece of attribute information, it sends out a Leave message and starts this timer. If a Join message does not arrive before the timer expires, the information is de-registered. The Leave timer must be greater than or equal to 3x the Join timer. The Dell Networking OS default is **600ms**.
- **LeaveAll** — After startup, a GARP device globally starts a LeaveAll timer. After expiration of this interval, it sends out a LeaveAll message so that other GARP devices can re-register all relevant attribute information. The device then restarts the LeaveAll timer to begin a new cycle. The LeaveAll timer must be greater than or equal to 5x of the Leave timer. The Dell Networking OS default is **10000ms**.

Example of the `garp timer` Command

```
Dell(conf)#garp timer leav 1000
Dell(conf)#garp timers leave-all 5000
Dell(conf)#garp timer join 300

Verification:

Dell(conf)#do show garp timer
GARP Timers Value (milliseconds)
----------------------------------------
Join Timer      300
Leave Timer     1000
```

```
LeaveAll Timer  5000
Dell(conf)#
```

Dell Networking OS displays this message if an attempt is made to configure an invalid GARP timer:

```
Dell(conf)#garp timers join 300 % Error: Leave timer should be >= 3*Join timer.
```

18

Internet Group Management Protocol (IGMP)

Internet group management protocol (IGMP) is supported on the S6000 platform.

Multicast is premised on identifying many hosts by a single destination IP address; hosts represented by the same IP address are a multicast group. IGMP is a Layer 3 multicast protocol that hosts use to join or leave a multicast group. Multicast routing protocols (such as protocol-independent multicast [PIM]) use the information in IGMP messages to discover which groups are active and to populate the multicast routing table.

IGMP Implementation Information

- Dell Networking Operating System (OS) supports IGMP versions 1, 2, and 3 based on RFCs 1112, 2236, and 3376, respectively.
- Dell Networking OS does not support IGMP version 3 and versions 1 or 2 on the same subnet.
- Dell Networking systems cannot serve as an IGMP host or an IGMP version 1 IGMP Querier.
- Dell Networking OS automatically enables IGMP on interfaces on which you enable a multicast routing protocol.

IGMP Protocol Overview

IGMP has three versions. Version 3 obsoletes and is backwards-compatible with version 2; version 2 obsoletes version 1.

IGMP Version 2

IGMP version 2 improves on version 1 by specifying IGMP Leave messages, which allows hosts to notify routers that they no longer care about traffic for a particular group.

Leave messages reduce the amount of time that the router takes to stop forwarding traffic for a group to a subnet (leave latency) after the last host leaves the group. In version 1 hosts quietly leave groups, and the router waits for a query response timer several times the value of the query interval to expire before it stops forwarding traffic.

To receive multicast traffic from a particular source, a host must join the multicast group to which the source is sending traffic. A host that is a member of a group is called a receiver. A host may join many groups, and may join or leave any group at any time. A host joins and leaves a multicast group by sending an IGMP message to its IGMP Querier. The querier is the router that surveys a subnet for multicast receivers and processes survey responses to populate the multicast routing table.

IGMP messages are encapsulated in IP packets, as shown in the following illustration.

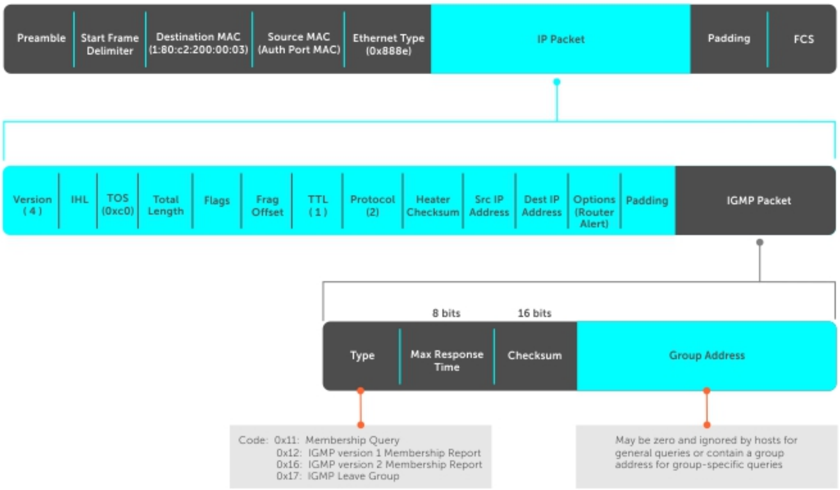

Preamble
Start Frame Delimiter
Destination MAC (1:80:c2:200:00:03)
Source MAC (Auth Port MAC)
Ethernet Type (0x888e)
IP Packet
Padding
FCS
Version (4)
IHL
TOS (0xc0)
Total Length
Flags
Frag Offset
TTL (1)
Protocol (2)
Heater Checksum
Src IP Address
Dest IP Address
Options (Router Alert)
Padding
IGMP Packet
8 bits
16 bits
Type
Max Response Time
Checksum
Group Address
Code: 0x11: Membership Query
0x12: IGMP version 1 Membership Report
0x16: IGMP version 2 Membership Report
0x17: IGMP Leave Group
May be zero and ignored by hosts for general queries or contain a group address for group-specific queries

Figure 37. IGMP Messages in IP Packets

Join a Multicast Group

There are two ways that a host may join a multicast group: it may respond to a general query from its querier or it may send an unsolicited report to its querier.

Responding to an IGMP Query

The following describes how a host can join a multicast group.

1. One router on a subnet is elected as the querier. The querier periodically multicasts (to all-multicast-systems address 224.0.0.1) a general query to all hosts on the subnet.
2. A host that wants to join a multicast group responds with an IGMP Membership Report that contains the multicast address of the group it wants to join (the packet is addressed to the same group). If multiple hosts want to join the same multicast group, only the report from the first host to respond reaches the querier and the remaining hosts suppress their responses (For how the delay timer mechanism works, refer to Adjusting Query and Response Timers).
3. The querier receives the report for a group and adds the group to the list of multicast groups associated with its outgoing port to the subnet. Multicast traffic for the group is then forwarded to that subnet.

Sending an Unsolicited IGMP Report

A host does not have to wait for a general query to join a group. It may send an unsolicited IGMP Membership Report, also called an IGMP Join message, to the querier.

Leaving a Multicast Group

The following describes how a host can leave a multicast group.

1. A host sends a membership report of type 0x17 (IGMP Leave message) to the all routers multicast address 224.0.0.2 when it no longer cares about multicast traffic for a particular group.
2. The querier sends a Group-Specific Query to determine whether there are any remaining hosts in the group. There must be at least one receiver in a group on a subnet for a router to forward multicast traffic for that group to the subnet.
3. Any remaining hosts respond to the query according to the delay timer mechanism (refer to Adjusting Query and Response Timers). If no hosts respond (because there are none remaining in the group), the querier waits a specified period and sends another query. If it still receives no

response, the querier removes the group from the list associated with forwarding port and stops forwarding traffic for that group to the subnet.

IGMP Version 3

Conceptually, IGMP version 3 behaves the same as version 2. However, there are differences.

- Version 3 adds the ability to filter by multicast source, which helps multicast routing protocols avoid forwarding traffic to subnets where there are no interested receivers.
- To enable filtering, routers must keep track of more state information, that is, the list of sources that must be filtered. An additional query type, the Group-and-Source-Specific Query, keeps track of state changes, while the Group-Specific and General queries still refresh the existing state.
- Reporting is more efficient and robust: hosts do not suppress query responses (non-suppression helps track state and enables the immediate-leave and IGMP snooping features), state-change reports are retransmitted to insure delivery, and a single membership report bundles multiple statements from a single host, rather than sending an individual packet for each statement.

The version 3 packet structure is different from version 2 to accommodate these protocol enhancements. Queries are still sent to the all-systems address 224.0.0.1, as shown in the following illustration, but reports are sent to the all IGMP version 3-capable multicast routers address 244.0.0.22, as shown in the second illustration.

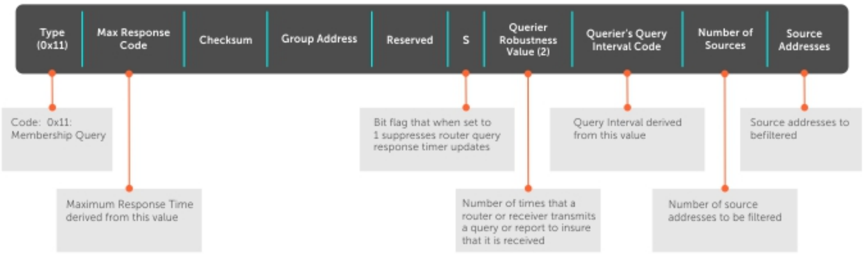
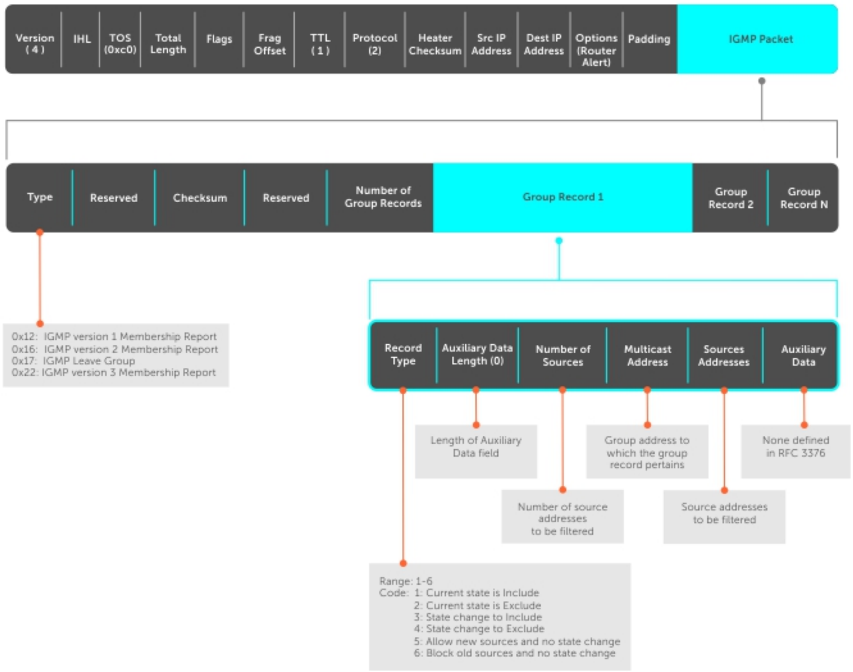

Type (0x11)
Max Response Code
Checksum
Group Address
Reserved
S
Querier Robustness Value (2)
Querier's Query Interval Code
Number of Sources
Source Addresses
Code: 0x11: Membership Query
Maximum Response Time derived from this value
Bit flag that when set to 1 suppresses router query response timer updates
Number of times that a router or receiver transmits a query or report to insure that it is received
Query Interval derived from this value
Number of source addresses to be filtered
Source addresses to befiltered

Figure 38. IGMP Version 3 Packet Structure



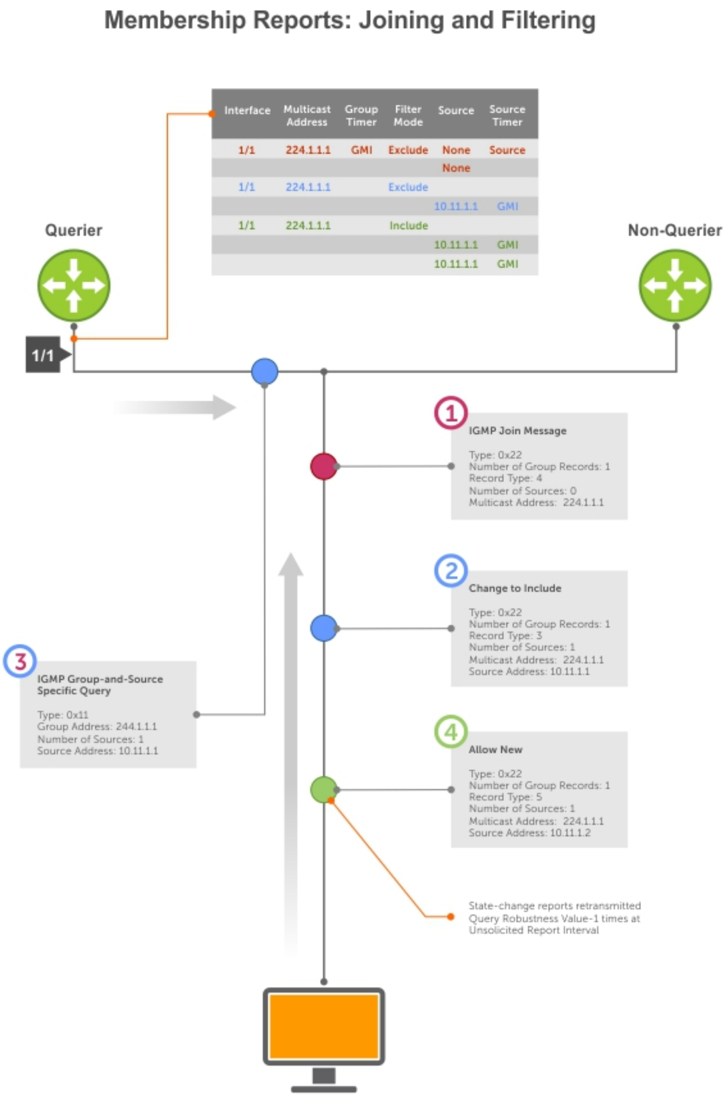
Version (4)
IHL
TOS (0xc0)
Total Length
Flags
Frag Offset
TTL (1)
Protocol (2)
Heater Checksum
Src IP Address
Dest IP Address
Options (Router Alert)
Padding
IGMP Packet
Type
Reserved
Checksum
Reserved
Number of Group Records
Group Record 1
Group Record 2
Group Record N
0x12: IGMP version 1 Membership Report
0x16: IGMP version 2 Membership Report
0x17: IGMP Leave Group
0x22: IGMP version 3 Membership Report
Record Type
Auxiliary Data Length (0)
Number of Sources
Multicast Address
Sources Addresses
Auxiliary Data
Length of Auxiliary Data field
Group address to which the group record pertains
None defined in RFC 3376
Number of source addresses to be filtered
Source addresses to be filtered
Range: 1-6
Code: 1: Current state is Include
2: Current state is Exclude
3: State change to Include
4: State change to Exclude
5: Allow new sources and no state change
6: Block old sources and no state change

Figure 39. IGMP Version 3–Capable Multicast Routers Address Structure

Joining and Filtering Groups and Sources

The following illustration shows how multicast routers maintain the group and source information from unsolicited reports.

1. The first unsolicited report from the host indicates that it wants to receive traffic for group 224.1.1.1.
2. The host’s second report indicates that it is only interested in traffic from group 224.1.1.1, source 10.11.1.1. Include messages prevents traffic from all other sources in the group from reaching the subnet. Before recording this request, the querier sends a group-and-source query to verify that there are no hosts interested in any other sources. The multicast router must satisfy all hosts if they have conflicting requests. For example, if another host on the subnet is interested in traffic from 10.11.1.3, the router cannot record the include request. There are no other interested hosts, so the request is recorded. At this point, the multicast routing protocol prunes the tree to all but the specified sources.
3. The host’s third message indicates that it is only interested in traffic from sources 10.11.1.1 and 10.11.1.2. Because this request again prevents all other sources from reaching the subnet, the router sends another group-and-source query so that it can satisfy all other hosts. There are no other interested hosts so the request is recorded.



Interface	Multicast Address	Group Timer	Filter Mode	Source	Source Timer
1/1	224.1.1.1	GMI	Exclude	None	Source
				None	
1/1	224.1.1.1		Exclude		
				10.11.1.1	GMI
1/1	224.1.1.1		Include		
				10.11.1.1	GMI
				10.11.1.1	GMI

① **IGMP Join Message**

Type: 0x22
Number of Group Records: 1
Record Type: 4
Number of Sources: 0
Multicast Address: 224.1.1.1

② **Change to Include**

Type: 0x22
Number of Group Records: 1
Record Type: 3
Number of Sources: 1
Multicast Address: 224.1.1.1
Source Address: 10.11.1.1

③ **IGMP Group-and-Source Specific Query**

Type: 0x11
Group Address: 244.1.1.1
Number of Sources: 1
Source Address: 10.11.1.1

④ **Allow New**

Type: 0x22
Number of Group Records: 1
Record Type: 5
Number of Sources: 1
Multicast Address: 224.1.1.1
Source Address: 10.11.1.2

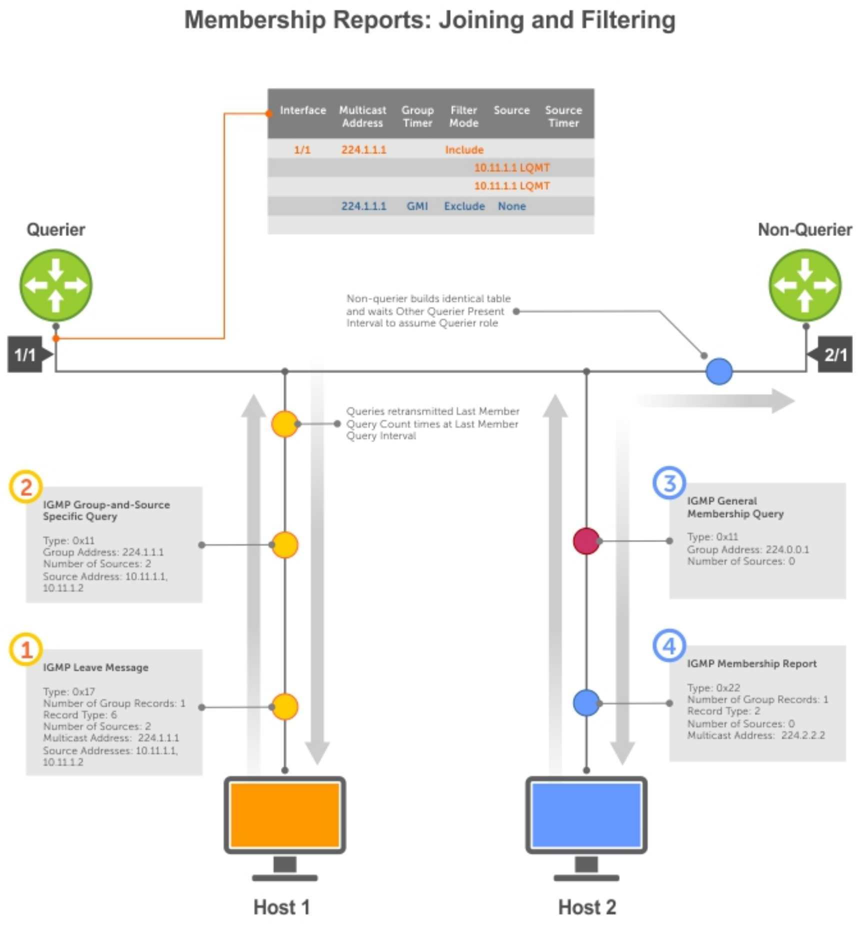
State-change reports retransmitted Query Robustness Value-1 times at Unsolicited Report Interval

Figure 40. Membership Reports: Joining and Filtering

Leaving and Staying in Groups

The following illustration shows how multicast routers track and refresh state changes in response to group-and-specific and general queries.

1. Host 1 sends a message indicating it is leaving group 224.1.1.1 and that the included filter for 10.11.1.1 and 10.11.1.2 are no longer necessary.
2. The querier, before making any state changes, sends a group-and-source query to see if any other host is interested in these two sources; queries for state-changes are retransmitted multiple times. If any are, they respond with their current state information and the querier refreshes the relevant state information.
3. Separately in the following illustration, the querier sends a general query to 224.0.0.1.
4. Host 2 responds to the periodic general query so the querier refreshes the state information for that group.


Membership Reports: Joining and Filtering
Querier
Non-Querier
1/1
2/1
Non-querier builds identical table and waits Other Querier Present Interval to assume Querier role
Queries retransmitted Last Member Query Count times at Last Member Query Interval
2
1
3
4
Host 1
Host 2

Interface	Multicast Address	Group Timer	Filter Mode	Source	Source Timer
1/1	224.1.1.1		Include		
				10.11.1.1 LQMT	
				10.11.1.1 LQMT	
	224.1.1.1	GMI	Exclude	None	

IGMP Group-and-Source Specific Query

Type: 0x11
Group Address: 224.1.1.1
Number of Sources: 2
Source Address: 10.11.1.1, 10.11.1.2

IGMP Leave Message

Type: 0x17
Number of Group Records: 1
Record Type: 6
Number of Sources: 2
Multicast Address: 224.1.1.1
Source Addresses: 10.11.1.1, 10.11.1.2

IGMP General Membership Query

Type: 0x11
Group Address: 224.0.0.1
Number of Sources: 0

IGMP Membership Report

Type: 0x22
Number of Group Records: 1
Record Type: 2
Number of Sources: 0
Multicast Address: 224.2.2.2

Figure 41. Membership Queries: Leaving and Staying

Configure IGMP

Configuring IGMP is a two-step process.

1. Enable multicast routing using the `ip multicast-routing` command.
2. Enable a multicast routing protocol.

Related Configuration Tasks

- Viewing IGMP Enabled Interfaces
- Selecting an IGMP Version
- Viewing IGMP Groups
- Adjusting Timers
- Configuring a Static IGMP Group
- Preventing a Host from Joining a Group
- Enabling IGMP Immediate-Leave
- IGMP Snooping

- Fast Convergence after MSTP Topology Changes
- Designating a Multicast Router Interface

Viewing IGMP Enabled Interfaces

Interfaces that are enabled with PIM-SM are automatically enabled with IGMP.
To view IGMP-enabled interfaces, use the following command.

- View IGMP-enabled interfaces.

  EXEC Privilege mode

  ```
  show ip igmp interface
  ```

Example of the `show ip igmp interface` Command

```
Dell#show ip igmp interface gig 7/16
GigabitEthernet 7/16 is up, line protocol is up
  Internet address is 10.87.3.2/24
  IGMP is enabled on interface
  IGMP query interval is 60 seconds
  IGMP querier timeout is 300 seconds
  IGMP max query response time is 10 seconds
  Last member query response interval is 199 ms
  IGMP activity: 0 joins, 0 leaves
  IGMP querying router is 10.87.3.2 (this system)
  IGMP version is 2
Dell#
```

Selecting an IGMP Version

Dell Networking OS enables IGMP version 2 by default, which supports version 1 and 2 hosts, but is not compatible with version 3 on the same subnet.
If hosts require IGMP version 3, you can switch to IGMP version 3.

To switch to version 3, use the following command.

- Switch to a different IGMP version.

  INTERFACE mode

  ```
  ip igmp version
  ```

Example of the `ip igmp version` Command

```
Dell(conf-if-gi-1/13)#ip igmp version 3
Dell(conf-if-gi-1/13)#do show ip igmp interface
GigabitEthernet 1/13 is up, line protocol is down
  Inbound IGMP access group is not set
  Interface IGMP group join rate limit is not set
  Internet address is 1.1.1.1/24
  IGMP is enabled on interface
  IGMP query interval is 60 seconds
  IGMP querier timeout is 125 seconds
  IGMP max query response time is 10 seconds
  IGMP last member query response interval is 1000 ms
  IGMP immediate-leave is disabled
  IGMP activity: 0 joins, 0 leaves, 0 channel joins, 0 channel leaves
  IGMP querying router is 1.1.1.1 (this system)
```

IGMP version is 3

```
Dell(conf-if-gi-1/13)#
```

Viewing IGMP Groups

To view both learned and statically configured IGMP groups, use the following command.

- View both learned and statically configured IGMP groups.

  EXEC Privilege mode

  ```
  show ip igmp groups
  ```

Example of the `show ip igmp groups` Command

```
Dell(conf-if-gi-1/0)#do sho ip igmp groups
Total Number of Groups: 2
IGMP Connected Group Membership
Group Address Interface          Uptime    Expires   Last Reporter
224.1.1.1     GigabitEthernet 1/0  00:00:03  Never     CLI
224.1.2.1     GigabitEthernet 1/0  00:56:55  00:01:22  1.1.1.2
```

Adjusting Timers

The following sections describe viewing and adjusting timers.
To view the current value of all IGMP timers, use the following command.

- View the current value of all IGMP timers.

  EXEC Privilege mode

  ```
  show ip igmp interface
  ```

For more information, refer to the example shown in Viewing IGMP Enabled Interfaces.

Adjusting Query and Response Timers

The querier periodically sends a general query to discover which multicast groups are active. A group must have at least one host to be active.
When a host receives a query, it does not respond immediately, but rather starts a delay timer. The delay time is set to a random value between 0 and the maximum response time. The host sends a response when the timer expires; in version 2, if another host responds before the timer expires, the timer is nullified, and no response is sent.

The maximum response time is the amount of time that the querier waits for a response to a query before taking further action. The querier advertises this value in the query (refer to the illustration in IGMP Version 2). Lowering this value decreases leave latency but increases response burstiness because all host membership reports must be sent before the maximum response time expires. Inversely, increasing this value decreases burstiness at the expense of leave latency.

When the querier receives a leave message from a host, it sends a group-specific query to the subnet. If no response is received, it sends another. The amount of time that the querier waits to receive a response to the initial query before sending a second one is the last member query interval (LMQI). The switch waits one LMQI after the second query before removing the group from the state table.

- Adjust the period between queries.

INTERFACE mode

```
ip igmp query-interval
```

- Adjust the maximum response time.

  INTERFACE mode

  ```
  ip igmp query-max-resp-time
  ```

- Adjust the last member query interval.

  INTERFACE mode

  ```
  ip igmp last-member-query-interval
  ```

Adjusting the IGMP Querier Timeout Value

If there is more than one multicast router on a subnet, only one is elected to be the querier, which is the router that sends queries to the subnet.

1. Routers send queries to the all multicast systems address, 224.0.0.1. Initially, all routers send queries.
2. When a router receives a query, it compares the IP address of the interface on which it was received with the source IP address given in the query. If the receiving router IP address is greater than the source address given in the query, the router stops sending queries. By this method, the router with the lowest IP address on the subnet is elected querier and continues to send queries.
3. If a specified amount of time elapses during which other routers on the subnet do not receive a query, those routers assume that the querier is down and a new querier is elected.

The amount of time that elapses before routers on a subnet assume that the querier is down is the other querier present interval.

- Adjust the other querier present interval.

  INTERFACE mode

  ```
  ip igmp querier-timeout
  ```

Configuring a Static IGMP Group

To configure and view a static IGMP group, use the following commands.
Multicast traffic for static groups is always forwarded to the subnet even if there are no members in the group.

Static groups have an expiration value of *Never* and a Last Reporter value of *CLI*, as shown in the example in Viewing IGMP Groups.

- Configure a static IGMP group.

  INTERFACE mode

  ```
  ip igmp static-group
  ```

- View the static groups.

  EXEC Privilege mode.

  ```
  show ip igmp groups
  ```

Enabling IGMP Immediate-Leave

If the querier does not receive a response to a group-specific or group-and-source query, it sends another (querier robustness value). Then, after no response, it removes the group from the outgoing interface for the subnet.
IGMP immediate leave reduces leave latency by enabling a router to immediately delete the group membership on an interface after receiving a Leave message (it does not send any group-specific or group-and-source queries before deleting the entry).

- Configure the system for IGMP immediate leave.

  ```
  ip igmp immediate-leave
  ```
- View the enable status of the IGMP immediate leave feature.

  EXEC Privilege mode

  ```
  show ip igmp interface
  ```

View the enable status of this feature using the command from EXEC Privilege mode, as shown in the example in Selecting an IGMP Version.

IGMP Snooping

IGMP snooping enables switches to use information in IGMP packets to generate a forwarding table that associates ports with multicast groups so that when they receive multicast frames, they can forward them only to interested receivers.

Multicast packets are addressed with multicast MAC addresses, which represent a group of devices, rather than one unique device. Switches forward multicast frames out of all ports in a virtual local area network (VLAN) by default, even though there may be only some interested hosts, which is a waste of bandwidth.

If you enable IGMP snooping on a VLT unit, IGMP snooping dynamically learned groups and multicast router ports are made to learn on the peer by explicitly tunneling the received IGMP control packets.

IGMP Snooping Implementation Information

- IGMP snooping on Dell Networking OS uses IP multicast addresses not MAC addresses.
- IGMP snooping reacts to spanning tree protocol (STP) and multiple spanning tree protocol (MSTP) topology changes by sending a general query on the interface that transitions to the forwarding state.
- If IGMP snooping is enabled on a PIM-enabled VLAN interface, data packets using the router as an Layer 2 hop may be dropped. To avoid this scenario, Dell Networking recommends that users enable IGMP snooping on server-facing end-point VLANs only.

Configuring IGMP Snooping

Configuring IGMP snooping is a one-step process. To enable, view, or disable IGMP snooping, use the following commands.
There is no specific configuration needed for IGMP snooping with virtual link trunking (VLT). For information about VLT configurations, refer to Virtual Link Trunking (VLT).

- Enable IGMP snooping on a switch.

  CONFIGURATION mode

  ```
  ip igmp snooping enable
  ```

- View the configuration.

  CONFIGURATION mode

  ```
  show running-config
  ```
- Disable snooping on a VLAN.

  INTERFACE VLAN mode

  ```
  no ip igmp snooping
  ```

Related Configuration Tasks

- Removing a Group-Port Association
- Disabling Multicast Flooding
- Specifying a Port as Connected to a Multicast Router
- Configuring the Switch as Querier

Example of `ip igmp snooping enable` Command

```
Dell(conf)#ip igmp snooping enable
Dell(conf)#do show running-config igmp
ip igmp snooping enable
Dell(conf)#
```

Removing a Group-Port Association

To configure or view the remove a group-port association feature, use the following commands.

- Configure the switch to remove a group-port association after receiving an IGMP Leave message.

  INTERFACE VLAN mode

  ```
  ip igmp fast-leave
  ```
- View the configuration.

  INTERFACE VLAN mode

  ```
  show config
  ```

Example of Configuration Output After Removing a Group-Port Association

```
Dell(conf-if-vl-100)#show config
!
interface Vlan 100
  no ip address
  ip igmp snooping fast-leave
  shutdown
Dell(conf-if-vl-100)#
```

Disabling Multicast Flooding

If the switch receives a multicast packet that has an IP address of a group it has not learned (unregistered frame), the switch floods that packet out of all ports on the VLAN.

When you configure the `no ip igmp snooping flood` command, the system drops the packets immediately. The system does not forward the frames on mrouter ports, even if they are present. Disable Layer 3 multicast (`no ip multicast-routing`) in order to disable multicast flooding.

- Configure the switch to only forward unregistered packets to ports on a VLAN that are connected to mrouter ports.

  CONFIGURATION mode

  ```
  no ip igmp snooping flood
  ```

Specifying a Port as Connected to a Multicast Router

To statically specify or view a port in a VLAN, use the following commands.

- Statically specify a port in a VLAN as connected to a multicast router.

  INTERFACE VLAN mode

  ```
  ip igmp snooping mrouter
  ```
- View the ports that are connected to multicast routers.

  EXEC Privilege mode.

  ```
  show ip igmp snooping mrouter
  ```

Configuring the Switch as Querier

To configure the switch as a querier, use the following command.
Hosts that do not support unsolicited reporting wait for a general query before sending a membership report. When the multicast source and receivers are in the same VLAN, multicast traffic is not routed and so there is no querier. Configure the switch to be the querier for a VLAN so that hosts send membership reports and the switch can generate a forwarding table by snooping.

- Configure the switch to be the querier for a VLAN by first assigning an IP address to the VLAN interface.

  INTERFACE VLAN mode

  ```
  ip igmp snooping querier
  ```

  IGMP snooping querier does not start if there is a statically configured multicast router interface in the VLAN.

  The switch may lose the querier election if it does not have the lowest IP address of all potential queriers on the subnet.

  When enabled, IGMP snooping querier starts after one query interval in case no IGMP general query (with IP SA lower than its VLAN IP address) is received on any of its VLAN members.

Adjusting the Last Member Query Interval

To adjust the last member query interval, use the following command.
When the querier receives a Leave message from a receiver, it sends a group-specific query out of the ports specified in the forwarding table. If no response is received, it sends another. The amount of time that the querier waits to receive a response to the initial query before sending a second one is the last member query interval (LMQI). The switch waits one LMQI after the second query before removing the group-port entry from the forwarding table.

- Adjust the last member query interval.

  INTERFACE VLAN mode

```
ip igmp snooping last-member-query-interval
```

Fast Convergence after MSTP Topology Changes

The following describes the fast convergence feature.

When a port transitions to the Forwarding state as a result of an STP or MSTP topology change, Dell Networking OS sends a general query out of all ports except the multicast router ports. The host sends a response to the general query and the forwarding database is updated without having to wait for the query interval to expire.

When an IGMP snooping switch is not acting as a querier, it sends out the general query in response to the MSTP triggered link-layer topology change, with the source IP address of 0.0.0.0 to avoid triggering querier election.

Egress Interface Selection (EIS) for HTTP and IGMP Applications

You can use the Egress Interface Selection (EIS) feature to isolate the management and front-end port domains for HTTP and IGMP traffic. Also, EIS enables you to configure the responses to switch-destined traffic by using the management port IP address as the source IP address. This information is sent out of the switch through the management port instead of the front-end port.

The management EIS feature is applicable only for the out-of-band (OOB) management port. References in this section to the management default route or static route denote the routes configured using the `management route` command. The management default route can be either configured statically or returned dynamically by the DHCP client. A static route points to the management interface or a forwarding router.

Transit traffic (destination IP not configured in the switch) that is received on the front-end port with destination on the management port is dropped and received in the management port with destination on the front-end port is dropped.

Switch-destined traffic (destination IP configured in the switch) is:

- Received in the front-end port with destination IP equal to management port IP address or management port subnet broadcast address is dropped.
- Received in the management port with destination IP not equal to management IP address or management subnet broadcast address is dropped.

Traffic (switch initiated management traffic or responses to switch-destined traffic with management port IP address as the source IP address) for user-specified management protocols must exit out of the management port. In this chapter, all the references to traffic indicate switch-initiated traffic and responses to switch-destined traffic with management port IP address as the source IP address.

In customer deployment topologies, it might be required that the traffic for certain management applications needs to exit out of the management port only. You can use EIS to control and the traffic can exit out of any port based on the route lookup in the IP stack.

One typical example is an SSH session to an unknown destination or an SSH connection that is destined to the management port IP address. The management default route can coexist with front-end default routes. If SSH is specified as a management application, SSH links to and from an unknown destination uses the management default route.

Protocol Separation

When you configure the `application` *`application-type`* command to configure a set of management applications with TCP/UDP port numbers to the OS, the following table describes the association between applications and their port numbers.

Table 27. Association Between Applications and Port Numbers

Application Name	Port Number	Client	Server
SSH	22	Supported	Supported
Sflow-Collector	6343	Supported	
SNMP	162 for SNMP Traps (client), 161 for SNMP MIB response (server)	Supported	
NTP	123	Supported	
DNS	53	Supported	
FTP	20/21	Supported	Supported
Syslog	514	Supported	
Telnet	23	Supported	Supported
TFTP	69	Supported	
Radius	1812,1813	Supported	
Tacacs	49	Supported	
HTTP	80 for httpd 443 for secure httpd 8008 HTTP server port for confd application 8888 secure HTTP server port for confd application		Supported

If you configure a source interface is for any EIS management application, EIS might not coexist with that interface and the behavior is undefined in such a case. You can configure the source interface for the following applications: FTP, ICMP (ping and traceroute utilities), NTP, RADIUS, TACACS, Telnet, TFTP, syslog, and SNMP traps. Out of these applications, EIS can coexist with only syslog and SNMP traps because these applications do not require a response after a packet is sent.

The switch also processes user-specified port numbers for applications such as RADIUS, TACACS, SSH, and sFlow. The OS maintains a list of configured management applications and their port numbers. You can configure two default routes, one configured on the management port and the other on the front-end port.

Two tables, namely, Egress Interface Selection routing table and default routing table, are maintained. In the preceding table, the columns Client and Server indicate that the applications can act as both a client and a server within the switch. The Management Egress Interface Selection table contains all management routes (connected, static and default route). The default routing table contains all management routes (connected, static and default route) and all front-end port routes.

Enabling and Disabling Management Egress Interface Selection

You can enable or disable egress-interface-selection using the `management egress-interface-selection` command.

NOTE: Egress Interface Selection (EIS) works only with IPv4 routing.

When the feature is enabled using the `management egress-interface-selection` command, the following events are performed:

- The CLI prompt changes to the EIS mode.
- In this mode, you can run the `application` and `no application` commands
- Applications can be configured or unconfigured as management applications using the `application` or `no application` command. All configured applications are considered as management applications and the rest of them as non-management applications.
- All the management routes (connected, static and default) are duplicated and added to the management EIS routing table.
- Any new management route added is installed to both the EIS routing table and default routing table.
- For management applications, route lookup is preferentially done in the management EIS routing table for all traffic. management port is the preferred egress port. For example, if SSH is a management application, an SSH session to a front-panel port IP on the peer box is initiated via management port only, if the management port is UP and management route is available.
- If SSH request is received on the management port destined to the management port IP address, the response to the request is sent out of the management port by performing a route lookup in the EIS routing table
- If the SSH request is received on the front-end port destined for the front-end IP address, the response traffic is sent by doing a route lookup in the default routing table only.
- If the management port is down or route lookup fails in the management EIS routing table, packets are dropped.
- For all non-management applications, traffic exits out of either front-end data port or management port based on route lookup in default routing table.
- Ping and traceroute are always non-management applications and route lookup for these applications is done in the default routing table only.
- For ping and traceroute utilities that are initiated from the switch, if reachability needs to be tested through routes in the management EIS routing table, you must configure ICMP as a management application.
- If ping and traceroute are destined to the management port IP address, the response traffic for these packets is sent by doing route lookup in the EIS routing table.

When the feature is disabled using the `no management egress-interface-selection` command, the following operations are performed:

- All management application configuration is removed.
- All routes installed in the management EIS routing table are removed.

Handling of Management Route Configuration

When the EIS feature is enabled, the following processing occurs:

- All existing management routes (connected, static and default) are duplicated and added to the management EIS routing table.
- Any management static route newly added using the `management route` CLI is installed to both the management EIS routing table and default routing table.
- As per existing behavior, for routes in the default routing table, conflicting front-end port routes if configured has higher precedence over management routes. So there can be scenarios where the same management route is present in the EIS routing table but not in the default routing table.
- Routes in the EIS routing table are displayed using the `show ip management-eis-route` command.
- In the netstat output, the prefix "mgmt" is added to routes in the EIS table so that the user can distinguish between routes in the EIS Routing table and default routing table.
- If the management port IP address is removed, the corresponding connected route is removed from both the EIS routing table and default routing table.
- If a management route is deleted, then the route is removed from both the EIS routing table and default routing table.

Handling of Switch-Initiated Traffic

When the control processor (CP) initiates a control packet, the following processing occurs:

- TCP/UDP port number is extracted from the sockaddr structure in the in_selectsrc call which is called as part of the connect system call or in the ip_output function. If the destination TCP/UDP port number belongs to a configured management application, then sin_port of destination sockaddr structure is set to Management EIS ID 2 so that route lookup can be done in the management EIS routing table.
- To ensure that protocol separation is done only for switch initiated traffic where the application acts as client, only the destination TCP/UDP port is compared and not the source TCP/UDP port. The source TCP/UDP port becomes a known port number when the box acts as server.
- TFTP is an exception to the preceding logic.
- For TFTP, data transfer is initiated on port 69, but the data transfer ports are chosen independently by the sender and receiver during initialization of the connection. The ports are chosen at random according to the parameters of the networking stack, typically from the range of temporary ports.
- If route lookup in EIS routing table succeeds, the application-specific packet count is incremented. This counter is viewed using the `show management application pkt-cntr` command. This counter is cleared using `clear management application pkt-cntr` command.
- If the route lookup in the EIS routing table fails or if management port is down, then packets are dropped. The application-specific count of the dropped packets is incremented and is viewed using

the `show management application pkt-drop-cntr` command. This counter is cleared using `clear management application pkt-drop-cntr` command.

- Packets whose destination TCP/UDP port does not match a configured management application, take the regular route lookup flow in the IP stack.
- In the ARP layer, for all ARP packets received through the management interface, a double route lookup is done, one in the default routing table and another in the management EIS routing table. This is because in the ARP layer, we do not have TCP/UDP port information to decide the table in which the route lookup should be done.
- The `show arp` command is enhanced to show the routing table type for the ARP entry.
- For the `clear arp-cache` command, upon receiving the ARP delete request, the route corresponding to the destination IP is identified. The ARP entries learned in the management EIS routing table are also cleared.
- Therefore, a separate control over clearing the ARP entries learned via routes in the EIS table is not present. If the ARP entry for a destination is cleared in the default routing table, then if an ARP entry for the destination exists in the EIS table, that entry is also cleared.
- Because fallback support is removed, if the management port is down or the route lookup in EIS table fails packets are dropped. Therefore, switch-initiated traffic sessions that used to work previously via fallback may not work now.

Handling of Switch-Destined Traffic

- The switch processes all traffic received on the management port destined to the management port IP address or the front-end port destined to the front-end IP address.
- If the source TCP/UDP port number matches a configured EIS or non-EIS management application and the source IP address is a management Port IP address, then the EIS route lookup is done for the response traffic and hence is sent out of the management port. In this case, the source IP address is a management port IP address only if the traffic was originally destined to the management port IP.
- ICMP-based applications like ping and traceroute are exceptions to the preceding logic since we do not have TCP/UDP port number. So if source IP address of the packet matches the management port IP address EIS route lookup is done.
- Management application packet counter is incremented if EIS route lookup succeeds and packet is sent out of the management port.
- If route lookup in the EIS routing table fails or if the management port is down, then packets are dropped. The management application drop counter is incremented.
- Whenever IP address is assigned to the management port, it is stored in a global variable in the IP stack, which is used for comparison with the source IP address of the packet.
- Rest of the response traffic is handled as per existing behavior by doing route lookup in the default routing table. So if the traffic is destined to the front-end port IP address, the response is sent out by doing a route lookup in the default routing table, which is an existing behavior.

Consider a sample topology in which ip1 is an address assigned to the management port and ip2 is an address assigned to any of the front panel port. A and B are end users on the management and front-panel port networks. The OS-initiated traffic for management applications takes a preference for ip1 as source IP and uses the management network to reach the destination. If the management port is down or the route lookup in EIS routing table fails, ip2 is the source IP and the front-panel port is used to reach the destination. The fallback route between the management and data networks is used in such a case. At any given time, end users can access Dell Networking OS applications using either ip1 or ip2. Return traffic for such end-user-originated sessions destined to management port ip1 is handled using the EIS route lookup.

Handling of Transit Traffic (Traffic Separation)

This is forwarded traffic where destination IP is not an IP address configured in the switch.

- Packets received on the management port with destination on the front-end port is dropped.
- Packets received on the front-end port with destination on the management port is dropped.
- A separate drop counter is incremented for this case. This counter is viewed using the `netstat` command, like all other IP layer counters.

Consider a scenario in which ip1 is an address assigned to the management port and ip2 is an address assigned to any of the front panel port of a switch. End users on the management and front panel port networks are connected. In such an environment, traffic received in the management port destined on the data port network is dropped and traffic received in the front-end port destined on the management network is dropped.

Mapping of Management Applications and Traffic Type

The following table summarizes the behavior of applications for various types of traffic when the management egress interface selection feature is enabled.

Table 28. Mapping of Management Applications and Traffic Type

Traffic type / Application type	Switch initiated traffic	Switch-destined traffic	Transit Traffic
EIS Management Application	Management is the preferred egress port selected based on route lookup in EIS table. If the management port is down or the route lookup fails, packets are dropped.	If source TCP/UDP port matches a management application and source IP address is management port IP address, management port is the preferred egress port selected based on route lookup in EIS table. If management port is down or route lookup fails, packets are dropped	Traffic from management port to data port and from data port to management port is blocked
Non-EIS management application	Front-end default route will take higher precedence over management default route and SSH session to an unknown destination uses the front-end default route only. No change in the existing behavior.	If source TCP/UDP port matches a management application and the source IP address is a management port IP address, the management port is the preferred egress port selected based on route lookup in EIS table. If the management port is down or the route lookup fails, packets are dropped	Traffic from management port to data port and from data port to management port is blocked

- EIS is enabled implies that EIS feature is enabled and the application might or might not be configured as a management application
- EIS is disabled implies that either EIS feature itself is disabled or that the application is not configured as a management application

Transit Traffic

This phenomenon occurs where traffic is transiting the switch. Traffic has not originated from the switch and is not terminating on the switch.

- Drop the packets that are received on the front-end data port with destination on the management port.
- Drop the packets that received on the management port with destination as the front-end data port.

Switch-Destined Traffic

This phenomenon occurs where traffic is terminated on the switch. Traffic has not originated from the switch and is not transiting the switch.

The switch accepts all traffic destined to the switch, which is received on management or front-end data port. Response traffic with management port IP address as source IP address is handled in the same manner as switch originated traffic.

Switch-Originated Traffic

This phenomenon occurs where traffic is originating from the switch.

1. Management Applications (Applications that are configured as management applications):

   The management port is an egress port for management applications. If the management port is down or the destination is not reachable through the management port (next hop ARP is not resolved, and so on), and if the destination is reachable through a data port, then the management application traffic is sent out through the front-end data port. This fallback mechanism is required.
2. Non-Management Applications (Applications that are not configured as management applications as defined by this feature):

   Non-management application traffic exits out of either front-end data port or management port based on routing table. If there is a default route on both the management and front-end data port, the default for the data port is preferred route.

Behavior of Various Applications for Switch-Initiated Traffic

This section describes the different system behaviors that occur when traffic is originating from the switch:

EIS Behavior: If the destination TCP/UDP port matches a configured management application, a route lookup is done in the EIS table and the management port gets selected as the egress port. If management port is down or the route lookup fails, packets are dropped.

EIS Behavior for ICMP: ICMP packets do not have TCP/UDP ports. To do an EIS route lookup for ICMP-based applications (ping and traceroute) using the source ip option, the management port IP address should be specified as the source IP address. If management port is down or route lookup fails, packets are dropped.

Default Behavior: Route lookup is done in the default routing table and appropriate egress port is selected.

Protocol	Behavior when EIS is Enabled	Behavior when EIS is Disabled
dns	EIS Behavior	Default Behavior
ftp	EIS Behavior	Default Behavior
ntp	EIS Behavior	Default Behavior
radius	EIS Behavior	Default Behavior
Sflow-collector		Default Behavior
Snmp (SNMP Mib response and SNMP Traps)	EIS Behavior	Default Behavior
ssh	EIS Behavior	Default Behavior
syslog	EIS Behavior	Default Behavior
tacacs	EIS Behavior	Default Behavior
telnet	EIS Behavior	Default Behavior
tftp	EIS Behavior	Default Behavior
icmp (ping and traceroute)	EIS Behavior for ICMP	Default Behavior

Behavior of Various Applications for Switch-Destined Traffic

This section describes the different system behaviors that occur when traffic is terminated on the switch. Traffic has not originated from the switch and is not transiting the switch. Switch-destined traffic is applicable only for applications which act as server for the TCP session and also for ICMP-based applications like ping and traceroute. FTP, SSH, and Telnet are the applications that can function as servers for the TCP session.

EIS Behavior: If source TCP or UDP port matches an EIS management or a non-EIS management application and source IP address is management port IP address, management port is the preferred egress port selected based on route lookup in EIS table. If the management port is down or the route lookup fails, packets are dropped.

If the source TCP/UDP port or source IP address does not match the management port IP address, a route lookup is done in the default routing table.

EIS behavior for ICMP: ICMP packets do not have TCP/UDP ports. In this case, to perform an EIS route lookup for ICMP-based applications (ping and traceroute), you must configure ICMP as a management application. If the management port is down or the route lookup fails, packets are dropped.

If source IP address does not match the management port IP address route lookup is done in the default routing table.

Default Behavior: Route lookup is done in the default routing table and appropriate egress port is selected.

Protocol	Behavior when EIS is Enabled	Behavior when EIS is Disabled
ftp	EIS Behavior	Default Behavior
http	EIS Behavior	Default Behavior
ssh	EIS Behavior	Default Behavior
Snmp (snmp mib response)	EIS Behavior	Default Behavior
telnet	EIS Behavior	Default Behavior
icmp (ping and traceroute)	EIS Behavior for ICMP	Default Behavior

Interworking of EIS With Various Applications

Stacking

- The management EIS is enabled on the master and the standby unit.
- Because traffic can be initiated from the Master unit only, the preference to management EIS table for switch-initiated traffic and all its related ARP processing is done in the Master unit only.
- ARP-related processing for switch-destined traffic is done by both master and standby units.

VLT

VLT feature is for the front-end port only. Because this feature is specific to the management port, this feature can coexist with VLT and nothing specific needs to be done in this feature to handle VLT scenario.

DHCP

- If DHCP Client is enabled on the management port, a management default route is installed to the switch.
- If management EIS is enabled, this default route is added to the management EIS routing table and the default routing table.

ARP learn enable

- When ARP learn enable is enabled, the switch learns ARP entries for ARP Request packets even if the packet is not destined to an IP configured in the box.
- The ARP learn enable feature is not applicable to the EIS routing table. It is applicable to the default routing table only to avoid unnecessary double ARP entries

Sflow

sFlow management application is supported only in standalone boxes and switch shall throw error message if sFlow is configured in stacking environment

Designating a Multicast Router Interface

To designate an interface as a multicast router interface, use the following command.
Dell Networking OS also has the capability of listening in on the incoming IGMP general queries and designate those interfaces as the multicast router interface when the frames have a non-zero IP source address. All IGMP control packets and IP multicast data traffic originating from receivers is forwarded to multicast router interfaces.

- Designate an interface as a multicast router interface.

  ```
  ip igmp snooping mrouter interface
  ```

19

Interfaces

This chapter describes interface types, both physical and logical, and how to configure them with Dell Networking Operating System (OS).

- 10 Gigabit Ethernet / 40 Gigabit Ethernet interfaces are supported on the S6000 platform.

Basic Interface Configuration

- Interface Types
- View Basic Interface Information
- Enabling a Physical Interface
- Physical Interfaces
- Management Interfaces
- VLAN Interfaces
- Loopback Interfaces
- Null Interfaces
- Port Channel Interfaces

Advanced Interface Configuration

- Bulk Configuration
- Defining Interface Range Macros
- Monitoring and Maintaining Interfaces
- Splitting QSFP Ports to SFP+ Ports
- Link Dampening
- Link Bundle Monitoring
- Ethernet Pause Frames
- Configure the MTU Size on an Interface
- Port-pipes
- Auto-Negotiation on Ethernet Interfaces
- View Advanced Interface Information

Interface Types

The following table describes different interface types.

Interface Type	Modes Possible	Default Mode	Requires Creation	Default State
Physical	L2, L3	Unset	No	Shutdown (disabled)
Management	N/A	N/A	No	No Shutdown (enabled)
Loopback	L3	L3	Yes	No Shutdown (enabled)
Null	N/A	N/A	No	Enabled
Port Channel	L2, L3	L3	Yes	Shutdown (disabled)
VLAN	L2, L3	L2	Yes (except default)	L2 - Shutdown (disabled) L3 - No Shutdown (enabled)

View Basic Interface Information

To view basic interface information, use the following command.
You have several options for viewing interface status and configuration parameters.

- Lists all configurable interfaces on the chassis.

  EXEC mode

  ```
  show interfaces
  ```

  This command has options to display the interface status, IP and MAC addresses, and multiple counters for the amount and type of traffic passing through the interface.

  If you configured a port channel interface, this command lists the interfaces configured in the port channel.

  NOTE: To end output from the system, such as the output from the `show interfaces` command, enter `CTRL+C` and Dell Networking OS returns to the command prompt.

  NOTE: The CLI output may be incorrectly displayed as 0 (zero) for the Rx/Tx power values. To obtain the correct power information, perform a simple network management protocol (SNMP) query.

Examples of the `show` Commands

The following example shows the configuration and status information for one interface.

```
Dell#show interfaces tengigabitethernet 1/0
TenGigabitEthernet 1/0 is up, line protocol is up
```

```
Hardware is Force10Eth, address is 00:01:e8:05:f3:6a
  Current address is 00:01:e8:05:f3:6a
Pluggable media present, XFP type is 10GBASE-LR.
  Medium is MultiRate, Wavelength is 1310nm
  XFP receive power reading is -3.7685
Interface index is 67436603
Internet address is 65.113.24.238/28
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 10000 Mbit, Mode full duplex, Master
ARP type: ARPA, ARP Timeout 04:00:00
Last clearing of "show interface" counters 00:09:54
Queueing strategy: fifo
Input Statistics:
  0 packets, 0 bytes
  0 Vlans
  0 64-byte pkts, 0 over 64-byte pkts, 0 over 127-byte pkts
  0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
  0 Multicasts, 0 Broadcasts
  0 runts, 0 giants, 0 throttles
  0 CRC, 0 overrun, 0 discarded
Output Statistics:
  3 packets, 192 bytes, 0 underruns
  3 64-byte pkts, 0 over 64-byte pkts, 0 over 127-byte pkts
  0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
  0 Multicasts, 3 Broadcasts, 0 Unicasts
  0 Vlans, 0 throttles, 0 discarded, 0 collisions
Rate info (interval 299 seconds):
  Input 00.00 Mbits/sec, 0 packets/sec, 0.00% of line-rate
  Output 00.00 Mbits/sec, 0 packets/sec, 0.00% of line-rate
Time since last interface status change: 00:00:31
Dell#
```

To view which interfaces are enabled for Layer 3 data transmission, use the `show ip interfaces brief` command in EXEC Privilege mode. In the following example, GigabitEthernet interface 1/5 is in Layer 3 mode because an IP address has been assigned to it and the interface’s status is operationally up.

```
Dell#show ip interface brief
Interface            IP-Address  OK? Method  Status                 Protocol
GigabitEthernet 1/0  unassigned  NO  Manual  administratively down  down
GigabitEthernet 1/1  unassigned  NO  Manual  administratively down  down
GigabitEthernet 1/2  unassigned  YES Manual  up                     up
GigabitEthernet 1/3  unassigned  YES Manual  up                     up
GigabitEthernet 1/4  unassigned  YES Manual  up                     up
GigabitEthernet 1/5  10.10.10.1  YES Manual  up                     up
GigabitEthernet 1/6  unassigned  NO  Manual  administratively down  down
GigabitEthernet 1/7  unassigned  NO  Manual  administratively down  down
GigabitEthernet 1/8  unassigned  NO  Manual  administratively down  down
```

To view only configured interfaces, use the `show interfaces configured` command in the EXEC Privilege mode. In the previous example, GigabitEthernet interface 1/5 is in Layer 3 mode because an IP address has been assigned to it and the interface’s status is operationally up.

To determine which physical interfaces are available, use the `show running-config` command in EXEC mode. This command displays all physical interfaces available on the line cards.

```
Dell#show running
Current Configuration ...
!
interface GigabitEthernet 9/6
  no ip address
  shutdown
!
```

```
interface GigabitEthernet 9/7
  no ip address
  shutdown
!
interface GigabitEthernet 9/8
  no ip address
  shutdown
!
interface GigabitEthernet 9/9
  no ip address
  shutdown
```

Enabling a Physical Interface

After determining the type of physical interfaces available, to enable and configure the interfaces, enter INTERFACE mode by using the `interface interface slot/port` command.

1. Enter the keyword `interface` then the type of interface and slot/port information.

   CONFIGURATION mode

   ```
   interface interface
   ```

   - For the Management interface on the RPM, enter the keyword `ManagementEthernet` then the slot/port information.
   - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
   - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.
2. Enable the interface.

   INTERFACE mode

   ```
   no shutdown
   ```

To confirm that the interface is enabled, use the `show config` command in INTERFACE mode. To leave INTERFACE mode, use the `exit` command or `end` command. You cannot delete a physical interface.

Physical Interfaces

The *Management Ethernet interface* is a single RJ-45 Fast Ethernet port on each unit of the S6000

The interface provides dedicated management access to the system.

Stack—unit interfaces support Layer 2 and Layer 3 traffic over the 10/100/1000 and 10-Gigabit Ethernet interfaces. Synchronous optical network technologies interfaces with point-to-point protocol (PPP) encapsulation support Layer 3 traffic. These interfaces can also become part of virtual interfaces such as virtual local area networks (VLANs) or port channels.

For more information about VLANs, refer to Bulk Configuration. For more information on port channels, refer to Port Channel Interfaces.

Dell Networking OS Behavior: The S6000 system uses a single MAC address for all physical interfaces.

Configuration Task List for Physical Interfaces

By default, all interfaces are operationally disabled and traffic does not pass through them.

The following section includes information about optional configurations for physical interfaces:

- Overview of Layer Modes
- Configuring Layer 2 (Data Link) Mode
- Configuring Layer 2 (Interface) Mode
- Management Interfaces
- Auto-Negotiation on Ethernet Interfaces
- Adjusting the Keepalive Timer
- Clearing Interface Counters

40G to 1G Breakout Cable Adaptor

As a default config command, you can see the speed CLI with options, 10M, 100M and, 1000M for all front end interfaces. By default, the “show config” does not show the default speed. But if you change the speed to 100 and comeback to speed 1000, it will show in “show config” or if you OIR 10G port with speed 1000 with 40gto1g cable it will show you the default speed.
You can use speed command only to configure these on Mgmt-optic ports alone. Without any optic, if you configure the speed, the configured will be assigned as the port speed to support Provisioning via BMP.

User viewable Logs:

LOGS for optic insertion and removal are same as QSFP optics. You can use ‘'`show inventory media`’to check the optic type, or ‘'`show interface transceiver`'to know the optic properties.

show inventory media

```
   0     88     QSFP        4x1000BASE-T        US0XJYD04162059          Yes
   0     89     QSFP        4x1000BASE-T        US0XJYD04162059          Yes
   0     90     QSFP        4x1000BASE-T        US0XJYD04162059          Yes
   0     91     QSFP        4x1000BASE-T        US0XJYD04162059
Yes
```

show interface transceiver

```
QSFP 0 Serial ID Base Fields
QSFP 0 Id                      = 0x0d
QSFP 0 Ext Id                  = 0x00
QSFP 0 Connector               = 0x0c
QSFP 0 Transceiver Code        = 0x04 0x00 0x00 0x00 0x00 0x00 0x00 0x00
QSFP 0 Encoding                = 0x05
QSFP 0 Length(SFM)     Km      = 0x00
QSFP 0 Length(OM3)     2m      = 0x32
QSFP 0 Length(OM2)     1m      = 0x00
QSFP 0 Length(OM1)     1m      = 0x00
QSFP 0 Length(Copper) 1m       = 0x00
QSFP 0 Vendor Rev              = 0
```

Overview of Layer Modes

On all systems running Dell Networking OS, you can place physical interfaces, port channels, and VLANs in Layer 2 mode or Layer 3 mode.

By default, VLANs are in Layer 2 mode.

Type of Interface	Possible Modes	Requires Creation	Default State
10/100/1000 Ethernet, 10 Gigabit Ethernet	Layer 2 Layer 3	No	Shutdown (disabled)
Management	N/A	No	Shutdown (disabled)
Loopback	Layer 3	Yes	No shutdown (enabled)
Null interface	N/A	No	Enabled
Port Channel	Layer 2 Layer 3	Yes	Shutdown (disabled)
VLAN	Layer 2 Layer 3	Yes, except for the default VLAN.	No shutdown (disabled for Layer 2) Shutdown (active for Layer 3)

Configuring Layer 2 (Data Link) Mode

Do not configure switching or Layer 2 protocols such as spanning tree protocol (STP) on an interface unless the interface has been set to Layer 2 mode.
To set Layer 2 data transmissions through an individual interface, use the following command.

- Enable Layer 2 data transmissions through an individual interface.

  INTERFACE mode

  ```
  switchport
  ```

Example of a Basic Layer 2 Interface Configuration

```
Dell(conf-if)#show config
!
interface Port-channel 1
  no ip address
  switchport
  no shutdown
Dell(conf-if)#
```

Configuring Layer 2 (Interface) Mode

To configure an interface in Layer 2 mode, use the following commands.

- Enable the interface.

  INTERFACE mode

  ```
  no shutdown
  ```
- Place the interface in Layer 2 (switching) mode.

  INTERFACE mode

  ```
  switchport
  ```

For information about enabling and configuring the Spanning Tree Protocol, refer to Spanning Tree Protocol (STP).

To view the interfaces in Layer 2 mode, use the `show interfaces switchport` command in EXEC mode.

Configuring Layer 3 (Network) Mode

When you assign an IP address to a physical interface, you place it in Layer 3 mode.
To enable Layer 3 mode on an individual interface, use the following commands. In all interface types except VLANs, the `shutdown` command prevents all traffic from passing through the interface. In VLANs, the `shutdown` command prevents Layer 3 traffic from passing through the interface. Layer 2 traffic is unaffected by the `shutdown` command. One of the interfaces in the system must be in Layer 3 mode before you configure or enter a Layer 3 protocol mode (for example, OSPF).

- Enable Layer 3 on an individual interface

  INTERFACE mode

  ```
  ip address
  ```
- Enable the interface.

  INTERFACE mode

  ```
  no shutdown
  ```

Example of Error Due to Issuing a Layer 3 Command on a Layer 2 Interface

If an interface is in the incorrect layer mode for a given command, an error message is displayed (shown in bold). In the following example, the `ip address` command triggered an error message because the interface is in Layer 2 mode and the `ip address` command is a Layer 3 command only.

```
Dell(conf-if)#show config
!
interface TenGigabitEthernet 1/2
  no ip address
  switchport
  no shutdown
Dell(conf-if)#ip address 10.10.1.1 /24
% Error: Port is in Layer 2 mode Gi 1/2.
Dell(conf-if)#
```

To determine the configuration of an interface, use the `show config` command in INTERFACE mode or the various `show interface` commands in EXEC mode.

Configuring Layer 3 (Interface) Mode

To assign an IP address, use the following commands.

- Enable the interface.

  INTERFACE mode

  ```
  no shutdown
  ```
- Configure a primary IP address and mask on the interface.

  INTERFACE mode

  ```
  ip address ip-address mask [secondary]
  ```

The *`ip-address`* must be in dotted-decimal format (A.B.C.D) and the mask must be in slash format (/xx).

Add the keyword `secondary` if the IP address is the interface’s backup IP address.

Example of the `show ip interface` Command

You can only configure one primary IP address per interface. You can configure up to 255 secondary IP addresses on a single interface.

To view all interfaces to see with an IP address assigned, use the `show ip interfaces brief` command in EXEC mode as shown in View Basic Interface Information.

To view IP information on an interface in Layer 3 mode, use the `show ip interface` command in EXEC Privilege mode.

```
Dell>show ip int vlan 58
Vlan 58 is up, line protocol is up
Internet address is 1.1.49.1/24
Broadcast address is 1.1.49.255
Address determined by config file
MTU is 1554 bytes
Inbound access list is not set
Proxy ARP is enabled
Split Horizon is enabled
Poison Reverse is disabled
ICMP redirects are not sent
ICMP unreachables are not sent
```

Egress Interface Selection (EIS)

Egress Interface Selection (EIS) is available on the platform.

EIS allows you to isolate the management and front-end port domains by preventing switch-initiated traffic routing between the two domains. This feature provides additional security by preventing flooding attacks on front-end ports. The following protocols support EIS: DNS, FTP, NTP, RADIUS, sFlow, SNMP, SSH, Syslog, TACACS, Telnet, and TFTP. This feature does not support sFlow on stacked units.

When you enable this feature, all management routes (connected, static, and default) are copied to the management EIS routing table. Use the `management route` command to add new management routes to the default and EIS routing tables. Use the `show ip management-eis-route` command to view the EIS routes.

Important Points to Remember

- Deleting a management route removes the route from both the EIS routing table and the default routing table.
- If the management port is down or route lookup fails in the management EIS routing table, the outgoing interface is selected based on route lookup from the default routing table.
- If a route in the EIS table conflicts with a front-end port route, the front-end port route has precedence.
- Due to protocol, ARP packets received through the management port create two ARP entries (one for the lookup in the EIS table and one for the default routing table).

Configuring EIS

EIS is compatible with the following protocols: DNS, FTP, NTP, RADIUS, sFlow, SNMP, SSH, Syslog, TACACS, Telnet, and TFTP.

To enable and configure EIS, use the following commands:

1. Enter EIS mode.

   CONFIGURATION mode

   ```
   management egress-interface-selection
   ```
2. Configure which applications uses EIS.

   EIS mode

   ```
   application {all | application-type}
   ```

   NOTE: If you configure SNMP as the management application for EIS and you add a default management route, when you perform an SNMP walk and check the debugging logs for the source and destination IPs, the SNMP agent uses the destination address of incoming SNMP packets as the source address for outgoing SNMP responses for security.

Management Interfaces

The system supports the Management Ethernet interface as well as the standard interface on any port. You can use either method to connect to the system.

Configuring Management Interfaces

The dedicated Management interface provides management access to the system on the S6000 platform.

You can configure this interface with Dell Networking OS, but the configuration options on this interface are limited. You cannot configure Gateway addresses and IP addresses if it appears in the main routing table of Dell Networking OS. In addition, proxy ARP is not supported on this interface.

To configure a management interface, use the following commands.

- Enter the slot and the port (0) to configure a Management interface.

  CONFIGURATION mode

  ```
  interface managementethernet interface
  ```

  The slot range is 0.

  The slot range is from 0 to 5.
- Configure an IP address and mask on a Management interface.

  INTERFACE mode

  ```
  ip address ip-address mask
  ```

  - *ip-address mask*: enter an address in dotted-decimal format (A.B.C.D). The mask must be in / prefix format (/x).

Configuring Management Interfaces on the S-Series

You can manage the S-Series from any port.
To configure an IP address for the port, use the following commands. There is no separate management routing table, so configure all routes in the IP routing table (the `ip route` command).

- Configure an IP address.

  INTERFACE mode

  ```
  ip address
  ```
- Enable the interface.

  INTERFACE mode

  ```
  no shutdown
  ```
- The interface is the management interface.

  INTEFACE mode

  ```
  description
  ```

Example of the `show interface` and `show ip route` Commands

To display the configuration for a given port, use the `show interface` command in EXEC Privilege mode, as shown in the following example. To display the routing table, use the `show ip route` command in EXEC Privilege mode.

```
Dell#show int gig 0/48
GigabitEthernet 0/48 is up, line protocol is up
Description: This is the Managment Interface
Hardware is Force10Eth, address is 00:01:e8:cc:cc:ce
  Current address is 00:01:e8:cc:cc:ce
Pluggable media not present
Interface index is 46449666
Internet address is 10.11.131.240/23
[output omitted]
Dell#show ip route
Codes: C - connected, S - static, R - RIP,
  B - BGP, IN - internal BGP, EX - external BGP,LO - Locally Originated,
  O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
  N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
  E2 - OSPF external type 2, i - IS-IS, L1 - IS-IS level-1,
  L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
  > - non-active route, + - summary route

Gateway of last resort is 10.11.131.254 to network 0.0.0.0

    Destination    Gateway                    Dist/Metric Last Change
    -----------    -------                    ----------- -----------
*S  0.0.0.0/0      via 10.11.131.254, Gi 0/48  1/0 1d2h
C  10.11.130.0/23  Direct, Gi 0/48             0/0 1d2h
Dell#
```

VLAN Interfaces

VLANs are logical interfaces and are, by default, in Layer 2 mode. Physical interfaces and port channels can be members of VLANs.
For more information about VLANs and Layer 2, refer to Layer 2 and Virtual LANs (VLANs).

NOTE: To monitor VLAN interfaces, use Management Information Base for Network Management of TCP/IP-based internets: MIB-II (RFC 1213).

NOTE: You cannot simultaneously use egress rate shaping and ingress rate policing on the same VLAN.

Dell Networking OS supports Inter-VLAN routing (Layer 3 routing in VLANs). You can add IP addresses to VLANs and use them in routing protocols in the same manner that physical interfaces are used. For more information about configuring different routing protocols, refer to the chapters on the specific protocol.

A consideration for including VLANs in routing protocols is that you must configure the `no shutdown` command. (For routing traffic to flow, you must enable the VLAN.)

NOTE: You cannot assign an IP address to the default VLAN, which is VLAN 1 (by default). To assign another VLAN ID to the default VLAN, use the `default vlan-id vlan-id` command.

To assign an IP address to an interface, use the following command.

- Configure an IP address and mask on the interface.
  INTERFACE mode

  ```
  ip address ip-address mask [secondary]
  ```

  - `ip-address mask`: enter an address in dotted-decimal format (A.B.C.D). The mask must be in slash format (/24).
  - `secondary`: the IP address is the interface's backup IP address. You can configure up to eight secondary IP addresses.

Example of a Configuration for a VLAN Participating in an OSPF Process

```
interface Vlan 10
  ip address 1.1.1.2/24
  tagged GigabitEthernet 2/2-13
  tagged TenGigabitEthernet 5/0
  ip ospf authentication-key force10
  ip ospf cost 1
  ip ospf dead-interval 60
  ip ospf hello-interval 15
  no shutdown
!
```

Loopback Interfaces

A Loopback interface is a virtual interface in which the software emulates an interface. Packets routed to it are processed locally.
Because this interface is not a physical interface, you can configure routing protocols on this interface to provide protocol stability. You can place Loopback interfaces in default Layer 3 mode.

To configure, view, or delete a Loopback interface, use the following commands.

- Enter a number as the Loopback interface.
  CONFIGURATION mode

  ```
  interface loopback number
  ```

  The range is from 0 to 16383.

- View Loopback interface configurations.

  EXEC mode

  ```
  show interface loopback number
  ```
- Delete a Loopback interface.

  CONFIGURATION mode

  ```
  no interface loopback number
  ```

Many of the same commands found in the physical interface are also found in the Loopback interfaces.

Null Interfaces

The Null interface is another virtual interface. There is only one Null interface. It is always up, but no traffic is transmitted through this interface.
To enter INTERFACE mode of the Null interface, use the following command.

- Enter INTERFACE mode of the Null interface.

  CONFIGURATION mode

  ```
  interface null 0
  ```

The only configurable command in INTERFACE mode of the Null interface is the `ip unreachable` command.

Port Channel Interfaces

Port channel interfaces support link aggregation, as described in IEEE Standard 802.3ad.

This section covers the following topics:

- Port Channel Definition and Standards
- Port Channel Benefits
- Port Channel Implementation
- Configuration Tasks for Port Channel Interfaces

Port Channel Definition and Standards

Link aggregation is defined by IEEE 802.3ad as a method of grouping multiple physical interfaces into a single logical interface—a link aggregation group (LAG) or port channel.

A LAG is “a group of links that appear to a MAC client as if they were a single link” according to IEEE 802.3ad. In Dell Networking OS, a LAG is referred to as a port channel interface.

A port channel provides redundancy by aggregating physical interfaces into one logical interface. If one physical interface goes down in the port channel, another physical interface carries the traffic.

Port Channel Benefits

A port channel interface provides many benefits, including easy management, link redundancy, and sharing.

Port channels are transparent to network configurations and can be modified and managed as one interface. For example, you configure one IP address for the group and that IP address is used for all routed traffic on the port channel.

With this feature, you can create larger-capacity interfaces by utilizing a group of lower-speed links. For example, you can build a 5-Gigabit interface by aggregating five 1-Gigabit Ethernet interfaces together. If one of the five interfaces fails, traffic is redistributed across the four remaining interfaces.

Port Channel Implementation

Dell Networking OS supports static and dynamic port channels.

- **Static** — Port channels that are statically configured.
- **Dynamic** — Port channels that are dynamically configured using the link aggregation control protocol (LACP). For details, refer to Link Aggregation Control Protocol (LACP).

There are 128 port-channels with eight members per channel.

NOTE: If you are using either 10G ports or 40G ports, the S6000 supports eight members per LAG.

As soon as you configure a port channel, Dell Networking OS treats it like a physical interface. For example, IEEE 802.1Q tagging is maintained while the physical interface is in the port channel.

Member ports of a LAG are added and programmed into the hardware in a predictable order based on the port ID, instead of in the order in which the ports come up. With this implementation, load balancing yields predictable results across line card resets and chassis reloads.

A physical interface can belong to only one port channel at a time.

Each port channel must contain interfaces of the same interface type/speed.

Port channels can contain a mix of 10, 100, or 1000 Mbps Ethernet interfaces and Gigabit Ethernet interfaces. The interface speed (10, 100, or 1000 Mbps) the port channel uses is determined by the first port channel member that is physically up. Dell Networking OS disables the interfaces that do match the interface speed that the first channel member sets. That first interface may be the first interface that is physically brought up or was physically operating when interfaces were added to the port channel. For example, if the first operational interface in the port channel is a Gigabit Ethernet interface, all interfaces at 1000 Mbps are kept up, and all 10/100/1000 interfaces that are not set to 1000 speed or auto negotiate are disabled.

Dell Networking OS brings up 10/100/1000 interfaces that are set to auto negotiate so that their speed is identical to the speed of the first channel member in the port channel.

10/100/1000 Mbps Interfaces in Port Channels

When both 10/100/1000 interfaces and GigE interfaces are added to a port channel, the interfaces must share a common speed. When interfaces have a configured speed different from the port channel speed, the software disables those interfaces.

The common speed is determined when the port channel is first enabled. At that time, the software checks the first interface listed in the port channel configuration. If you enabled that interface, its speed configuration becomes the common speed of the port channel. If the other interfaces configured in that port channel are configured with a different speed, Dell Networking OS disables them.

For example, if four interfaces (TenGig 0/1, 0/2, 0/3 and 0/4) in which TenGig 0/1 and TenGig 0/2 are set to speed 1000 Mb/s and the others(te 0/3 and 0/4) are set to 10000 Mb/s, with all interfaces enabled, and you add them to a port channel by entering `channel-member tengigabitethernet 0/1-4` while in port channel interface mode, and Dell Networking OS determines if the first interface specified (TenGig 0/1) is up. After it is up, the common speed of the port channel is 1000 Mb/s. Dell Networking OS disables those interfaces configured with speed 10000 Mb/s or whose speed is 10000 Mb/s as a result of auto-negotiation.

In this example, you can change the common speed of the port channel by changing its configuration so the first enabled interface referenced in the configuration is a 1000 Mb/s speed interface. You can also change the common speed of the port channel here by setting the speed of the Gi 0/0 interface to 1000 Mb/s.

Configuration Tasks for Port Channel Interfaces

To configure a port channel (LAG), use the commands similar to those found in physical interfaces. By default, no port channels are configured in the startup configuration.

These are the mandatory and optional configuration tasks:

- Creating a Port Channel (mandatory)
- Adding a Physical Interface to a Port Channel (mandatory)
- Reassigning an Interface to a New Port Channel (optional)
- Configuring the Minimum Oper Up Links in a Port Channel (optional)
- Adding or Removing a Port Channel from a VLAN (optional)
- Assigning an IP Address to a Port Channel (optional)
- Deleting or Disabling a Port Channel (optional)
- Load Balancing Through Port Channels (optional)

Creating a Port Channel

You can create up to 128 port channels with eight port members per group on the .
To configure a port channel, use the following commands.

1. Create a port channel.

   CONFIGURATION mode

   ```
   interface port-channel id-number
   ```
2. Ensure that the port channel is active.

   INTERFACE PORT-CHANNEL mode

   ```
   no shutdown
   ```

After you enable the port channel, you can place it in Layer 2 or Layer 3 mode. To place the port channel in Layer 2 mode or configure an IP address to place the port channel in Layer 3 mode, use the `switchport` command.

You can configure a port channel as you would a physical interface by enabling or configuring protocols or assigning access control lists.

Adding a Physical Interface to a Port Channel

The physical interfaces in a port channel can be on any line card in the chassis, but must be the same physical type.

NOTE: Port channels can contain a mix of Gigabit Ethernet and 10/100/1000 Ethernet interfaces, but Dell Networking OS disables the interfaces that are not the same speed of the first channel member in the port channel (refer to 10/100/1000 Mbps Interfaces in Port Channels).

You can add any physical interface to a port channel if the interface configuration is minimal. You can configure only the following commands on an interface if it is a member of a port channel:

- `description`
- `shutdown/no shutdown`
- `mtu`
- `ip mtu` (if the interface is on a Jumbo-enabled by default)

NOTE:

A logical port channel interface cannot have flow control. Flow control can only be present on the physical interfaces if they are part of a port channel.

NOTE: To configure the MTU, use the `mtu` command from INTERFACE mode.

To view the interface's configuration, enter INTERFACE mode for that interface and use the `show config` command or from EXEC Privilege mode, use the `show running-config interface` *`interface`* command.

When an interface is added to a port channel, Dell Networking OS recalculates the hash algorithm.

To add a physical interface to a port, use the following commands.

1. Add the interface to a port channel.
   INTERFACE PORT-CHANNEL mode

   ```
   channel-member interface
   ```

   The *`interface`* variable is the physical interface type and slot/port information.
2. Double check that the interface was added to the port channel.
   INTERFACE PORT-CHANNEL mode

   ```
   show config
   ```

Examples of the `show interfaces port-channel` Commands

To view the port channel's status and channel members in a tabular format, use the `show interfaces port-channel brief` command in EXEC Privilege mode, as shown in the following example.

```
Dell#show int port brief

LAG Mode Status Uptime   Ports
1 L2L3   up     00:06:03 Gi 13/6 (Up) *
                         Gi 13/12 (Up)
2 L2L3   up     00:06:03 Gi 13/7 (Up) *
```

```
                         Gi 13/8 (Up)
                         Gi 13/13 (Up)
                         Gi 13/14 (Up)
Dell#
```

The following example shows the port channel's mode (L2 for Layer 2 and L3 for Layer 3 and L2L3 for a Layer 2-port channel assigned to a routed VLAN), the status, and the number of interfaces belonging to the port channel.

```
Dell>show interface port-channel 20
Port-channel 20 is up, line protocol is up
Hardware address is 00:01:e8:01:46:fa
Internet address is 1.1.120.1/24
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 2000 Mbit
Members in this channel: Gi 9/10 Gi 9/17
ARP type: ARPA, ARP timeout 04:00:00
Last clearing of "show interface" counters 00:00:00
Queueing strategy: fifo
  1212627 packets input, 1539872850 bytes
  Input 1212448 IP Packets, 0 Vlans 0 MPLS
  4857 64-byte pkts, 17570 over 64-byte pkts, 35209 over 127-byte pkts
  69164 over 255-byte pkts, 143346 over 511-byte pkts, 942523 over 1023-byte
pkts
  Received 0 input symbol errors, 0 runts, 0 giants, 0 throttles
  42 CRC, 0 IP Checksum, 0 overrun, 0 discarded
  2456590833 packets output, 203958235255 bytes, 0 underruns
  Output 1640 Multicasts, 56612 Broadcasts, 2456532581 Unicasts
  2456590654 IP Packets, 0 Vlans, 0 MPLS
  0 throttles, 0 discarded
Rate info (interval 5 minutes):
  Input 00.01Mbits/sec, 2 packets/sec
  Output 81.60Mbits/sec, 133658 packets/sec
Time since last interface status change: 04:31:57

Dell>
```

When more than one interface is added to a Layer 2-port channel, Dell Networking OS selects one of the active interfaces in the port channel to be the primary port. The primary port replies to flooding and sends protocol data units (PDUs). An asterisk in the `show interfaces port-channel brief` command indicates the primary port.

As soon as a physical interface is added to a port channel, the properties of the port channel determine the properties of the physical interface. The configuration and status of the port channel are also applied to the physical interfaces within the port channel. For example, if the port channel is in Layer 2 mode, you cannot add an IP address or a static MAC address to an interface that is part of that port channel. In the following example, interface GigabitEthernet 1/6 is part of port channel 5, which is in Layer 2 mode, and an error message appeared when an IP address was configured.

```
Dell(conf-if-portch)#show config
!
interface Port-channel 5
  no ip address
  switchport
  channel-member GigabitEthernet 1/6
Dell(conf-if-portch)#int gi 1/6
Dell(conf-if)#ip address 10.56.4.4 /24
% Error: Port is part of a LAG Gi 1/6.
Dell(conf-if)#
```

Reassigning an Interface to a New Port Channel

An interface can be a member of only one port channel. If the interface is a member of a port channel, remove it from the first port channel and then add it to the second port channel.
Each time you add or remove a channel member from a port channel, Dell Networking OS recalculates the hash algorithm for the port channel.

To reassign an interface to a new port channel, use the following commands.

1. Remove the interface from the first port channel.

   INTERFACE PORT-CHANNEL mode

   `no channel-member` *`interface`*
2. Change to the second port channel INTERFACE mode.

   INTERFACE PORT-CHANNEL mode

   `interface port-channel id` *`number`*
3. Add the interface to the second port channel.

   INTERFACE PORT-CHANNEL mode

   `channel-member interface`

Example of Moving an Interface to a New Port Channel

The following example shows moving the TenGigabitEthernet 0/8 interface from port channel 4 to port channel 3.

```
Dell(conf-if-po-4)#show config
!
interface Port-channel 4
  no ip address
  channel-member TenGigabitEthernet 0/8
  no shutdown
Dell(conf-if-po-4)#no chann tengi 0/8
Dell(conf-if-po-4)#int port 3
Dell(conf-if-po-3)#channel tengi 0/8
Dell(conf-if-po-3)#sho conf
!
interface Port-channel 3
  no ip address
  channel-member TenGigabitEthernet 0/8
  shutdown
Dell(conf-if-po-3)#
```

Configuring the Minimum Oper Up Links in a Port Channel

You can configure the minimum links in a port channel (LAG) that must be in “oper up” status to consider the port channel to be in “oper up” status.
To set the “oper up” status of your links, use the following command.

- Enter the number of links in a LAG that must be in “oper up” status.

  INTERFACE mode

  `minimum-links` *`number`*

  The default is **1**.

Example of Configuring the Minimum Oper Up Links in a Port Channel

```
Dell#config t
Dell(conf)#int po 1
Dell(conf-if-po-1)#minimum-links 5
Dell(conf-if-po-1)#
```

Configuring VLAN Tags for Member Interfaces

To configure and verify VLAN tags for individual members of a port channel, perform the following:

1. Configure VLAN membership on individual ports

   INTERFACE mode

   ```
   Dell(conf-if-te-0/2)#vlan tagged 2,3-4
   ```
2. Use the `switchport` command in INTERFACE mode to enable Layer 2 data transmissions through an individual interface

   INTERFACE mode

   ```
   Dell(conf-if-te-0/2)#switchport
   ```
3. Verify the manually configured VLAN membership (**show interfaces switchport** *interface* command).

   EXEC mode

   ```
   Dell(conf)# interface tengigabitethernet 0/1
   Dell(conf-if-te-0/1)#switchport
   Dell(conf-if-te-0/1)# vlan tagged 2-5,100,4010
   Dell#show interfaces switchport te 0/1

   Codes:  U - Untagged, T - Tagged
           x - Dot1x untagged, X - Dot1x tagged
           G - GVRP tagged, M - Trunk, H - VSN tagged
           i - Internal untagged, I - Internal tagged, v - VLT untagged, V -
   VLT tagged

   Name: TenGigabitEthernet 0/1
   802.1QTagged: True
   Vlan membership:
   Q       Vlans
   T       2-5,100,4010

   Dell#
   ```

Assigning an IP Address to a Port Channel

You can assign an IP address to a port channel and use port channels in Layer 3 routing protocols.
To assign an IP address, use the following command.

- Configure an IP address and mask on the interface.

  INTERFACE mode

  ```
  ip address ip-address mask [secondary]
  ```

  - *`ip-address mask`*: enter an address in dotted-decimal format (A.B.C.D). The mask must be in slash format (/24).

  - `secondary`: the IP address is the interface's backup IP address. You can configure up to eight secondary IP addresses.

Deleting or Disabling a Port Channel

To delete or disable a port channel, use the following commands.

- Delete a port channel.

  CONFIGURATION mode

  ```
  no interface portchannel channel-number
  ```
- Disable a port channel.

  ```
  shutdown
  ```

  When you disable a port channel, all interfaces within the port channel are operationally down also.

Load Balancing Through Port Channels

Dell Networking OS uses hash algorithms for distributing traffic evenly over channel members in a port channel (LAG).

The hash algorithm distributes traffic among Equal Cost Multi-path (ECMP) paths and LAG members. The distribution is based on a flow, except for packet-based hashing. A flow is identified by the hash and is assigned to one link. In packet-based hashing, a single flow can be distributed on the LAG and uses one link.

Packet based hashing is used to load balance traffic across a port-channel based on the IP Identifier field within the packet. Load balancing uses source and destination packet information to get the greatest advantage of resources by distributing traffic over multiple paths when transferring data to a destination.

Dell Networking OS allows you to modify the hashing algorithms used for flows and for fragments. The load-balance and hash-algorithm commands are available for modifying the distribution algorithms.

NOTE: Hash-based load-balancing on multi-protocol label switching (MPLS) does not work when you enable packet-based hashing (`load-balance ip-selection packet-based`).

Load-Balancing on the S- Series

For LAG hashing on the S6000 the source IP, destination IP, source transmission control protocol (TCP)/ user datagram protocol (UDP) port, and destination TCP/UDP port are used for hash computation by default. For packets without a Layer 3 header, Dell Networking OS automatically uses `load-balance mac source-dest-mac`.

Do not configure IP hashing or MAC hashing at the same time. If you configure an IP and MAC hashing scheme at the same time, the MAC hashing scheme takes precedence over the IP hashing scheme.

To change the IP traffic load-balancing default, use the following command.

- Replace the default IP 4-tuple method of balancing traffic over a port channel.

  CONFIGURATION mode

  ```
  [no] load-balance {ip-selection [dest-ip | source-ip]} | {mac [dest-mac |
  source-dest-mac | source-mac]} | {tcp-udp enable} {ipv6-selection} {tunnel}|
  {ingress-port}
  ```

  You can select one, two, or all three of the following basic hash methods:

  - `ip-selection [dest-ip | source-ip]` — Distribute IP traffic based on the IP destination or source address.
  - `mac [dest-mac | source-dest-mac | source-mac]` — Distribute IPV4 traffic based on the destination or source MAC address, or both, along with the VLAN, Ethertype, source module ID and source port ID.
  - `tcp-udp enable` — Distribute traffic based on the TCP/UDP source and destination ports.
  - `ingress`-port — Option to Source Port Id for ECMP/ LAG hashing.
  - `ipv6-selection`— Set the IPV6 key fields to use in hash computation.
  - `tunnel`— Set the tunnel key fields to use in hash computation.

Changing the Hash Algorithm

The `load-balance` command selects the hash criteria applied to port channels.
If you do not obtain even distribution with the `load-balance` command, you can use the `hash-algorithm` command to select the hash scheme for LAG, ECMP and NH-ECMP. You can rotate or shift the 12–bit Lag Hash until the desired hash is achieved.

The `nh-ecmp` option allows you to change the hash value for recursive ECMP routes independently of non-recursive ECMP routes. This option provides for better traffic distribution over available equal cost links that involve a recursive next hop lookup.

To change to another algorithm, use the second command.

- Change the default (0) to another algorithm and apply it to ECMP, LAG hashing, or a particular line card.

  CONFIGURATION mode

  ```
  hash-algorithm | [ecmp{crc16|crc16cc|crc32LSB|crc32MSB|crc-upper|dest-ip |lsb
  |xor1| xor2| xor4| xor8| xor16}|lag{crc16|crc16cc|crc32LSB|crc32MSB|xor1|
  xor2|xor4|xor8|xor16}| seed ]
  ```

  For more information about algorithm choices, refer to the command details in the *IP Routing* chapter of the *Dell Networking OS Command Reference Guide*.
- Change the Hash algorithm seed value to get better hash value

  Hash seed is used to compute the hash value. By default hash seed is chassis MAC 32 bits. we can also change the hash seed by the following command.

  CONFIGURATION mode

  ```
  hash-algorithm seed {seed value}
  ```
- Change to another algorithm.

  CONFIGURATION mode

  ```
  hash-algorithm [ecmp{crc16|crc16cc|crc32LSB|crc32MSB|crc-upper|dest-ip|lsb|
  xor1|xor2|xor4|xor8|xor16}]
  ```

Example of the `hash-algorithm` Command

```
Dell(conf)#hash-algorithm ecmp xor 26 lag crc 26 nh-ecmp checksum 26
Dell(conf)#
```

The `hash-algorithm` command is specific to ECMP group. The default ECMP hash configuration is **crc-lower**. This command takes the lower 32 bits of the hash key to compute the egress port. Other options for ECMP hash-algorithms are:

- `crc16` — uses 16 bit CRC16-bisync polynomial
- `crc16cc` — uses 16 bit CRC16 using CRC16-CCITT polynomial
- `crc32LSB` — uses LSB 16 bits of computed CRC32
- `crc32MSB` — uses MSB 16 bits of computed CRC32(default)
- `crc-upper` — uses the upper 32 bits of the hash key to compute the egress port.
- `dest-ip` — uses destination IP address as part of the hash key.
- `lsb` —uses always return the LSB of the key as the hash
- `xor1` — uses Upper 8 bits of CRC16-BISYNC and lower 8 bits of xor1
- `xor2` — Upper 8 bits of CRC16-BISYNC and lower 8 bits of xor2
- `xor4` —Upper 8 bits of CRC16-BISYNC and lower 8 bits of xor4
- `xor8` — Upper 8 bits of CRC16-BISYNC and lower 8 bits of xor8
- `xor16` — uses 16 bit XOR.

Bulk Configuration

Bulk configuration allows you to determine if interfaces are present for physical interfaces or configured for logical interfaces.

Interface Range

An interface range is a set of interfaces to which other commands may be applied and may be created if there is at least one valid interface within the range.

Bulk configuration excludes from configuration any non-existing interfaces from an interface range. A default VLAN may be configured only if the interface range being configured consists of only VLAN ports.

The `interface range` command allows you to create an interface range allowing other commands to be applied to that range of interfaces.

The interface range prompt offers the interface (with slot and port information) for valid interfaces. The maximum size of an interface range prompt is 32. If the prompt size exceeds this maximum, it displays (...) at the end of the output.

NOTE: Non-existing interfaces are excluded from the interface range prompt. In the following example, Tengigabit 3/0 and VLAN 1000 do not exist.

NOTE: When creating an interface range, interfaces appear in the order they were entered and are not sorted.

The `show range` command is available under Interface Range mode. This command allows you to display all interfaces that have been validated under the interface range context.

The `show configuration` command is also available under Interface Range mode. This command allows you to display the running configuration only for interfaces that are part of interface range.

Bulk Configuration Examples

Use the `interface range` command for bulk configuration.

- Create a Single-Range
- Create a Multiple-Range
- Exclude Duplicate Entries
- Exclude a Smaller Port Range
- Overlap Port Ranges
- Commas
- Add Ranges

Create a Single-Range

The following is an example of a single range.

Example of the `interface range` Command (Single Range)

```
Dell(config)# interface range tengigabitethernet 0/1 - 23
Dell(config-if-range-tegi-0/1-23)# no shutdown
Dell(config-if-range-tegi-0/1-23)#
```

Create a Multiple-Range

The following is an example of multiple range.

Example of the `interface range` Command (Multiple Ranges)

```
Dell(conf)#interface range tengigabitethernet 0/5 - 10 , tengigabitethernet
0/1 , vlan 1
Dell(conf-if-range-te-0/5-10,te-0/1,vl-1)#
```

Exclude Duplicate Entries

The following is an example showing how duplicate entries are omitted from the interface-range prompt.

Example of the Interface-Range Prompt for Duplicate Interfaces

```
Dell(conf)#interface range vlan 1 , vlan 1 , vlan 3 , vlan 3
Dell(conf-if-range-vl-1,vl-3)#
Dell(conf)#interface range tengigabitethernet 2/0 - 23 , tengigabitethernet 2/0
- 23 , tengigab 2/0 - 23
Dell(conf-if-range-te-2/0-23)#
```

Exclude a Smaller Port Range

The following is an example show how the smaller of two port ranges is omitted in the interface-range prompt.

Example of the Interface-Range Prompt for Multiple Port Ranges

```
Dell(conf)#interface range tengigabitethernet 2/0 - 23 , tengigab 2/1 - 10
Dell(conf-if-range-te-2/0-23)#
```

Overlap Port Ranges

The following is an example showing how the interface-range prompt extends a port range from the smallest start port number to the largest end port number when port ranges overlap. handles overlapping port ranges.

Example of the Interface-Range Prompt for Overlapping Port Ranges

```
Dell(conf)#inte ra te 2/1 - 11 , te 2/1 - 23
Dell(conf-if-range-te-2/1-23)#
```

Commas

The following is an example of how to use commas to add different interface types to the range, enabling all Ten Gigabit Ethernet interfaces in the range 5/1 to 5/23 and both Ten Gigabit Ethernet interfaces 1/1 and 1/2.

Example of Adding Interface Ranges

```
Dell(config-if)# interface range tengigabitethernet 5/1 - 23,
tengigabitethernet 1/1 - 2
Dell(config-if-range-te-5/1-23)# no shutdown
Dell(config-if-range-te-5/1-23)#
```

Add Ranges

The following example shows how to use commas to add VLAN and port-channel interfaces to the range.

Example of Adding VLAN and Port-Channel Interface Ranges

```
Dell(config-if-range-te-1/1-2)# interface range Vlan 2 – 100 , Port 1 – 25
Dell(config-if-range-te-1/1-2-so-5/1-vl-2-100-po-1-25)# no shutdown
```

Defining Interface Range Macros

You can define an interface-range macro to automatically select a range of interfaces for configuration. Before you can use the `macro` keyword in the `interface-range macro` command string, define the macro.

To define an interface-range macro, use the following command.

- Defines the interface-range macro and saves it in the running configuration file.

  CONFIGURATION mode

  ```
  define interface-range macro_name {vlan vlan_ID - vlan_ID} |
  {{gigabitethernet | tengigabitethernet | fortyGigE} slot/interface -
  interface} [ , {vlan vlan_ID - vlan_ID} {{gigabitethernet |
  tengigabitethernet | fortyGigE} slot/interface - interface}]
  ```

Define the Interface Range

The following example shows how to define an interface-range macro named "test" to select Fast Ethernet interfaces 5/1 through 5/4.

Example of the `define interface-range` Command for Macros

```
Dell(config)# define interface-range test gigabitethernet 5/1 - 4
```

Choosing an Interface-Range Macro

To use an interface-range macro, use the following command.

- Selects the interfaces range to be configured using the values saved in a named interface-range macro.

  CONFIGURATION mode

  `interface range macro` *`name`*

Example of Using a Macro to Change the Interface Range Configuration Mode

The following example shows how to change to the interface-range configuration mode using the interface-range macro named “test.”

```
Dell(config)# interface range macro test
Dell(config-if)#
```

Monitoring and Maintaining Interfaces

Monitor interface statistics with the `monitor interface` command. This command displays an ongoing list of the interface status (up/down), number of packets, traffic statistics, and so on.
To view the interface’s statistics, use the following command.

- View the interface’s statistics.

  EXEC Privilege mode

  Enter the type of interface and slot/port information:

  - For the Management interface on the stack-unit, enter the keyword `ManagementEthernet` then the slot/port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
  - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.

Example of the `monitor interface` Command

The information displays in a continuous run, refreshing every 2 seconds by default. To manage the output, use the following keys.

- `m` — Change mode
- `l` — Page up
- `T` — Increase refresh interval (by 1 second)
- `t` — Decrease refresh interval (by 1 second)
- `c` — Clear screen
- `a` — Page down
- `q` — Quit

```
Dell#monitor interface Te 3/1

Dell uptime is 1 day(s), 4 hour(s), 31 minute(s)
  Monitor time: 00:00:00 Refresh Intvl.: 2s

Interface: TeGi 3/1, Disabled, Link is Down, Linespeed is 1000 Mbit
```

```
Traffic statistics: Current     Rate    Delta
       Input bytes:       0   0 Bps       0
      Output bytes:       0   0 Bps       0
     Input packets:       0   0 pps       0
    Output packets:       0   0 pps       0
       64B packets:       0   0 pps       0
  Over 64B packets:       0   0 pps       0
 Over 127B packets:       0   0 pps       0
 Over 255B packets:       0   0 pps       0
 Over 511B packets:       0   0 pps       0
Over 1023B packets:       0   0 pps       0
  Error statistics:
   Input underruns:       0   0 pps       0
      Input giants:       0   0 pps       0
   Input throttles:       0   0 pps       0
         Input CRC:       0   0 pps       0
 Input IP checksum:       0   0 pps       0
     Input overrun:       0   0 pps       0
  Output underruns:       0   0 pps       0
  Output throttles:       0   0 pps       0

m - Change mode                  c - Clear screen
l - Page up                      a - Page down
T - Increase refresh interval    t - Decrease refresh interval
q - Quit

q
Dell#
```

Maintenance Using TDR

The time domain reflectometer (TDR) is supported on all Dell Networking switch/routers.
TDR is an assistance tool to resolve link issues that helps detect obvious open or short conditions within any of the four copper pairs. TDR sends a signal onto the physical cable and examines the reflection of the signal that returns. By examining the reflection, TDR is able to indicate whether there is a cable fault (when the cable is broken, becomes unterminated, or if a transceiver is unplugged).

TDR is useful for troubleshooting an interface that is not establishing a link; that is, when the link is flapping or not coming up. TDR is not intended to be used on an interface that is passing traffic. When a TDR test is run on a physical cable, it is important to shut down the port on the far end of the cable. Otherwise, it may lead to incorrect test results.

NOTE: TDR is an intrusive test. Do not run TDR on a link that is up and passing traffic.

To test and display TDR results, use the following commands.

1. To test for cable faults on the GigabitEthernet cable.
   EXEC Privilege mode

   ```
   tdr-cable-test gigabitethernet <slot>/<port>
   ```

   Between two ports, do not start the test on both ends of the cable.

   Enable the interface before starting the test.

   Enable the port to run the test or the test prints an error message.
2. Displays TDR test results.

EXEC Privilege mode

```
show tdr gigabitethernet <slot>/<port>
```

Splitting QSFP Ports to SFP+ Ports

Splitting QSFP ports to SFP+ ports is supported on the S6000 platform.
The S6000 platform supports splitting a single 40G QSFP port into four 10G SFP+ ports using one of the supported breakout cables (for a list of supported cables, refer to the *Installation Guide* or the *Release Notes*).

NOTE: When you split a 40G port (such as fo 0/4) into four 10G ports, the 40G interface configuration is available in the startup configuration when you save the running configuration by using the write memory command. When a reload of the system occurs, the 40G interface configuration is not applicable because the 40G ports are split into four 10G ports after the reload operation. While the reload is in progress, you might see error messages when the configuration file is being loaded. You can ignore these error messages. Similarly, such error messages are displayed during a reload after you configure the four individual 10G ports to be stacked as a single 40G port.

To split a single 40G port into four 10G ports, use the following command.

- Split a single 40G port into 4-10G ports.

  CONFIGURATION mode

  ```
  stack-unit stack-unit port number portmode quad
  ```

  - `stack-unit`: enter the stack member unit identifier of the stack member to reset. The range is from 0 to 11.
  - `number`: enter the port number of the 40G port to be split.

Important Points to Remember

- Splitting a 40G port into four 10G ports is supported on standalone and stacked units.
- You cannot use split ports as stack-link to stack a system.
- The unit number with the split ports must be the default (**stack-unit 0**).

To verify port splitting, use the `show system brief` command. If the unit ID is different than 0, it must be renumbered to 0 before ports are split by using the `stackunit id renumber 0` command in EXEC mode.

- The quad port must be in a default configuration before you can split it into 4x10G ports. The 40G port is lost in the configuration when the port is split; be sure that the port is also removed from other L2/L3 feature configurations.
- The system must be reloaded after issuing the CLI for the change to take effect.

Converting a QSFP or QSFP+ Port to an SFP or SFP+ Port

You can convert a QSFP or QSFP+ port to an SFP or SFP+ port using the Quad to Small Form Factor Pluggable Adapter (QSA).

QSA provides smooth connectivity between devices that use Quad Lane Ports (such as the 40 Gigabit Ethernet adapters) and 10 Gigabit hardware that uses SFP+ based cabling. Using this adapter, you can effectively use a QSFP or QSFP+ module to connect to a lower-end switch or server that uses an SFP or SFP+ based module.

When connected to a QSFP or QSFP+ port on a 40 Gigabit adapter, QSA acts as an interface for the SFP or SFP+ cables. This interface enables you to directly plug in an SFP or SFP+ cable originating at a 10 Gigabit Ethernet port on a switch or server.

You can use QSFP optical cables (without a QSA) to split a 40 Gigabit port on a switch or a server into four 10 Gigabit ports. You must enable the fan-out mode in order for this mechanism to work. For more details, see Splitting QSFP Ports to SFP+ Ports.

Similarly, you can enable the fan-out mode to configure the QSFP port on a device to act as an SFP or SFP+ port. As the QSA enables a QSFP or QSFP+ port to be used as an SFP or SFP+ port, Dell Networking OS does not immediately detect the QSA after you insert it into a QSFP port cage.

After you insert an SFP or SFP+ cable into a QSA connected to a 40 Gigabit port, Dell Networking OS assumes that all the four fanned-out 10 Gigabit ports have plugged-in SFP or SFP+ optical cables. However, the link UP event happens only for the first 10 Gigabit port and you can use only that port for data transfer. As a result, only the first fanned-out port is identified as the active 10 Gigabit port with a speed of 10G or 1G depending on whether you insert an SFP+ or SFP cable respectively.

NOTE: Although it is possible to configure the remaining three 10 Gigabit ports, the Link UP event does not occur for these ports leaving the lanes unusable. Dell Networking OS perceives these ports to be in a Link Down state. You must not try to use these remaining three 10 Gigabit ports for actual data transfer or for any other related configurations.

NOTE: Trident2 chip sets do not work at 1G speeds with auto-negotiation enabled. As a result, when you peer any device using SFP, the link does not come up if auto-negotiation is enabled. Therefore, you must disable auto-negotiation on platforms that currently use Trident2 chip sets (S6000 and Z9000). This limitation applies only when you convert QSFP to SFP using the QSA. This constraint does not apply for QSFP to SFP+ conversions using the QSA.

Important Points to Remember

- Before using the QSA to convert a 40 Gigabit Ethernet port to a 10 Gigabit SFP or SFP+ port, you must enable 40 G to 4*10 fan-out mode on the device.
- When you insert a QSA into a 40 Gigabit port, you can use only the first 10 Gigabit port in the fan-out mode to plug-in SFP or SFP+ cables. The remaining three 10 Gigabit ports are perceived to be in Link Down state and are unusable.
- You cannot use QSFP optical cables in a QSA setup.
- When you remove the QSA module alone from a 40 Gigabit port, without connecting any SFP or SFP+ cables; Dell Networking OS does not generate any event. However, when you remove a QSA module that has SFP or SFP+ optical cables plugged in, Dell Networking OS generates a SFP or SFP+ Removed event.
- In the S6000 platform, you can use the QSA on any of the ports. However, the existing maximum fan-out restrictions apply to the ports.
- The QSA module does not have a designated EEPROM. To recognize a QSA, Dell Networking OS reads the EEPROM corresponding to a SFP+ or SFP module that is plugged into QSA. The access location of this EEPROM is different from the EEPROM location of the QSFP+ module.
- The diagnostics application is capable of detecting insertion or removal of both the QSA as well as the SFP+ or SFP optical cables plugged into the QSA. In addition, the diagnostic application is also capable of reading the DDS and Vendor information from the EEPROM corresponding to SFP+ or SFP optical cables. As a result, no separate detection of QSA is required.

Support for LM4 Optics

The newly supported LM4 optics are similar in behavior to the LR4 optics that are already supported. However, in the output of `show inventory media` command, an LM4 optical module is denoted as 40G-LM4. Barring this exception, the functionality and behavior of LM4 optics is similar to LR4 optics.

Example Scenarios

Consider the following scenarios:

- QSFP port 0 is connected to a QSA with SFP+ optical cables plugged in.
- QSFP port 4 is connected to a QSA with SFP optical cables plugged in.
- QSFP port 8 in fanned-out mode is plugged in with QSFP optical cables.
- QSFP port 12 in 40 G mode is plugged in with QSFP optical cables.

For these configurations, the following examples show the command output that the `show interfaces tengigbitethernet transceiver`, `show interfaces tengigbitethernet`, and `show inventory media` commands displays:

```
Dell#show interfaces tengigabitethernet 0/0 transceiver
SFP+ 0 Serial ID Base Fields
SFP+ 0 Id                       = 0x0d
SFP+ 0 Ext Id                   = 0x00
SFP+ 0 Connector                = 0x23
SFP+ 0 Transceiver Code         = 0x08 0x00 0x00 0x00 0x00 0x00 0x00 0x00
SFP+ 0 Encoding                 = 0x00
………………
………………
SFP+ 0 Diagnostic Information
===================================
SFP+ 0 Rx Power measurement type        = OMA
===================================
SFP+ 0 Temp High Alarm threshold        = 0.000C
SFP+ 0 Voltage High Alarm threshold     = 0.000V
SFP+ 0 Bias High Alarm threshold        = 0.000mA
```

NOTE: In the following `show interfaces tengigbitethernet` commands, the ports 1,2, and 3 are inactive and no physical SFP or SFP+ connection actually exists on these ports. However, Dell Networking OS still perceives these ports as valid and the output shows that pluggable media (optical cables) is inserted into these ports. This is a software limitation for this release.

```
Dell#show interfaces tengigabitethernet 0/1 transceiver
SFP+ 0 Serial ID Base Fields
SFP+ 0 Id                       = 0x0d
SFP+ 0 Ext Id                   = 0x00
SFP+ 0 Connector                = 0x23
…………………….

Dell#show interfaces tengigabitethernet 0/2 transceiver
SFP+ 0 Serial ID Base Fields
SFP+ 0 Id                       = 0x0d
SFP+ 0 Ext Id                   = 0x00
SFP+ 0 Connector                = 0x23
…………………….


Dell#show interfaces tengigabitethernet 0/3 transceiver
SFP+ 0 Serial ID Base Fields
```

```
SFP+ 0 Id                        = 0x0d
SFP+ 0 Ext Id                    = 0x00
SFP+ 0 Connector                 = 0x23
………………………….


Dell#show interfaces tengigabitethernet 0/4 transceiver
SFP 0 Serial ID Base Fields
SFP 0 Id                        = 0x0d
SFP 0 Ext Id                    = 0x00
SFP 0 Connector                 = 0x23
SFP 0 Transceiver Code          = 0x08 0x00 0x00 0x00 0x00 0x00 0x00 0x00
SFP 0 Encoding                  = 0x00
………………
………………
SFP 0 Diagnostic Information
===================================
SFP 0 Rx Power measurement type         = OMA
===================================
SFP 0 Temp High Alarm threshold         = 0.000C
SFP 0 Voltage High Alarm threshold      = 0.000V
SFP 0 Bias High Alarm threshold         = 0.000mA
```

NOTE: In the following `show interfaces tengigbitethernet transceiver` commands, the ports 5,6, and 7 are inactive and no physical SFP or SFP+ connection actually exists on these ports. However, Dell Networking OS still perceives these ports as valid and the output shows that pluggable media (optical cables) is inserted into these ports. This is a software limitation for this release.

```
Dell#show interfaces tengigabitethernet 0/5 transceiver
SFP 0 Serial ID Base Fields
SFP 0 Id                        = 0x0d
SFP 0 Ext Id                    = 0x00
SFP 0 Connector                 = 0x23
SFP 0 Transceiver Code          = 0x08 0x00 0x00 0x00 0x00 0x00 0x00 0x00
SFP 0 Encoding                  = 0x00
………………


Dell#show interfaces tengigabitethernet 0/6 transceiver
SFP 0 Serial ID Base Fields
SFP 0 Id                        = 0x0d
SFP 0 Ext Id                    = 0x00
SFP 0 Connector                 = 0x23
SFP 0 Transceiver Code          = 0x08 0x00 0x00 0x00 0x00 0x00 0x00 0x00
SFP 0 Encoding                  = 0x00
………………


Dell#show interfaces tengigabitethernet 0/7 transceiver
SFP 0 Serial ID Base Fields
SFP 0 Id                        = 0x0d
SFP 0 Ext Id                    = 0x00
SFP 0 Connector                 = 0x23
SFP 0 Transceiver Code          = 0x08 0x00 0x00 0x00 0x00 0x00 0x00 0x00
SFP 0 Encoding                  = 0x00
………………


Dell#show interfaces tengigabitethernet 0/8 transceiver
QSFP 0 Serial ID Base Fields
QSFP 0 Id                        = 0x0d
QSFP 0 Ext Id                    = 0x00
QSFP 0 Connector                = 0x23
QSFP 0 Transceiver Code         = 0x08 0x00 0x00 0x00 0x00 0x00 0x00 0x00
QSFP 0 Encoding                 = 0x00
………………
………………
QSFP 0 Diagnostic Information
===================================
QSFP 0 Rx Power measurement type        = OMA
===================================
QSFP 0 Temp High Alarm threshold        = 0.000C
QSFP 0 Voltage High Alarm threshold     = 0.000V
QSFP 0 Bias High Alarm threshold        = 0.000mA

Dell#show interfaces fortyGigE 0/12 transceiver
QSFP 0 Serial ID Base Fields
QSFP 0 Id                       = 0x0d
QSFP 0 Ext Id                   = 0x00
QSFP 0 Connector                = 0x23
QSFP 0 Transceiver Code         = 0x08 0x00 0x00 0x00 0x00 0x00 0x00 0x00
QSFP 0 Encoding                 = 0x00
………………
………………
QSFP 0 Diagnostic Information
===================================
QSFP 0 Rx Power measurement type        = OMA
===================================
QSFP 0 Temp High Alarm threshold        = 0.000C
QSFP 0 Voltage High Alarm threshold     = 0.000V
QSFP 0 Bias High Alarm threshold        = 0.000mA
$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$

Dell#show interfaces tengigabitethernet 0/0
tengigabitethernet 0/0 is up, line protocol is up
Hardware is DellEth, address is 90:b1:1c:f4:9a:fa
    Current address is 90:b1:1c:f4:9a:fa
Pluggable media present, SFP+ type is 10GBASE-SX
Interface index is 35012865
Internet address is not set
Mode of IPv4 Address Assignment : NONE
DHCP Client-ID :90b11cf49afa
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 10000 Mbit

Dell#show interfaces tengigabitethernet 0/1
tengigabitethernet 0/1 is up, line protocol is down
Hardware is DellEth, address is 90:b1:1c:f4:9a:fa
    Current address is 90:b1:1c:f4:9a:fa
Pluggable media present, SFP+ type is 10GBASE-SX
……….
LineSpeed 10000 Mbit

Dell#show interfaces tengigabitethernet 0/2
tengigabitethernet 0/1 is up, line protocol is down
Hardware is DellEth, address is 90:b1:1c:f4:9a:fa
    Current address is 90:b1:1c:f4:9a:fa
Pluggable media present, SFP+ type is 10GBASE-SX
……….
LineSpeed 10000 Mbit


Dell#show interfaces tengigabitethernet 0/3
tengigabitethernet 0/1 is up, line protocol is down
Hardware is DellEth, address is 90:b1:1c:f4:9a:fa
    Current address is 90:b1:1c:f4:9a:fa
```

```
Pluggable media present, SFP+ type is 10GBASE-SX
……….
LineSpeed 10000 Mbit

Dell#show interfaces tengigabitethernet 0/4
gigabitethernet 0/0 is up, line protocol is up
Hardware is DellEth, address is 90:b1:1c:f4:9a:fa
    Current address is 90:b1:1c:f4:9a:fa
Pluggable media present, SFP type is 1GBASE
……………………
LineSpeed 1000 Mbit

Dell#show interfaces tengigabitethernet 0/5
gigabitethernet 0/0 is up, line protocol is down
Hardware is DellEth, address is 90:b1:1c:f4:9a:fa
    Current address is 90:b1:1c:f4:9a:fa
Pluggable media present, SFP type is 1GBASE
……………………
LineSpeed 1000 Mbit

Dell#show interfaces tengigabitethernet 0/6
gigabitethernet 0/0 is up, line protocol is down
Hardware is DellEth, address is 90:b1:1c:f4:9a:fa
    Current address is 90:b1:1c:f4:9a:fa
Pluggable media present, SFP type is 1GBASE
……………………
LineSpeed 1000 Mbit

Dell#show interfaces tengigabitethernet 0/7
gigabitethernet 0/0 is up, line protocol is down
Hardware is DellEth, address is 90:b1:1c:f4:9a:fa
    Current address is 90:b1:1c:f4:9a:fa
Pluggable media present, SFP type is 1GBASE
……………………
LineSpeed 1000 Mbit

Dell#show interfaces tengigabitethernet 0/8
TenGigabitEthernet 0/0 is up, line protocol is up
Hardware is DellEth, address is 90:b1:1c:f4:9a:fa
    Current address is 90:b1:1c:f4:9a:fa
Pluggable media present, QSFP type is 4x10GBASE-CR1-3M
……..
LineSpeed 10000 Mbit
```

`The show inventory command shows the following output:`

NOTE: In the following `show inventory media` command output, the port numbers 1, 2, 3, 5, 6, and 7 ports are actually inactive. However, Dell Networking OS still shows that optical cables are inserted into these ports. This is a software limitation for this release.

```
Dell# show inventory media
Slot      Port     Type        Media                 Serial Number
-------------------------------------------------------------------
0          0       SFP+      10GBASE-SX            APF12420031B3P
0          1       SFP+      10GBASE-SX            APF12420031B3P
0          2       SFP+      10GBASE-SX            APF12420031B3P
0          3       SFP+      10GBASE-SX            APF12420031B3P
0          4       SFP       10GBASE-SX           APF12420031B3P
0          5       SFP       10GBASE-SX            APF12420031B3P
0          6       SFP       10GBASE-SX            APF12420031B3P
0          7       SFP       10GBASE-SX            APF12420031B3P
0          8       QSFP      4x10GBASE-CR1-3M      APF12420031B3P
0          9       QSFP      4x10GBASE-CR1-3M      APF12420031B3P
```

```
0          10       QSFP      4x10GBASE-CR1-3M        APF12420031B3P
0          11       QSFP      4x10GBASE-CR1-3M        APF12420031B3P
0          12       QSFP      40GBASE-SR4
```

Link Dampening

Interface state changes occur when interfaces are administratively brought up or down or if an interface state changes.

Every time an interface changes a state or flaps, routing protocols are notified of the status of the routes that are affected by the change in state. These protocols go through the momentous task of re-converging. Flapping; therefore, puts the status of entire network at risk of transient loops and black holes.

Link dampening minimizes the risk created by flapping by imposing a penalty for each interface flap and decaying the penalty exponentially. After the penalty exceeds a certain threshold, the interface is put in an Error-Disabled state and for all practical purposes of routing, the interface is deemed to be "down." After the interface becomes stable and the penalty decays below a certain threshold, the interface comes up again and the routing protocols re-converge.

Link dampening:

- reduces processing on the CPUs by reducing excessive interface flapping.
- improves network stability by penalizing misbehaving interfaces and redirecting traffic.
- improves convergence times and stability throughout the network by isolating failures so that disturbances are not propagated.

Important Points to Remember

- Link dampening is not supported on VLAN interfaces.
- Link dampening is disabled when the interface is configured for port monitoring.
- You can apply link dampening to Layer 2 and Layer 3 interfaces.
- You can configure link dampening on individual interfaces in a LAG.

Enabling Link Dampening

To enable link dampening, use the following command.

- Enable link dampening.

  INTERFACE mode

  ```
  dampening
  ```

Examples of the `show interfaces dampening` Commands

To view the link dampening configuration on an interface, use the `show config` command.

```
R1(conf-if-gi-1/1)#show config
!
interface GigabitEthernet 1/1
  ip address 10.10.19.1/24
  dampening 1 2 3 4
  no shutdown
R1(conf-if-gi-1/1)#exit
```

To view dampening information on all or specific dampened interfaces, use the `show interfaces dampening` command from EXEC Privilege mode.

```
Dell# show interfaces dampening
InterfaceStateFlapsPenaltyHalf-LifeReuseSuppressMax-Sup
Gi 0/0Up005750250020
Gi 0/1Up21200205001500300
Gi 0/2Down4850306002000120
```

To view a dampening summary for the entire system, use the `show interfaces dampening summary` command from EXEC Privilege mode.

```
Dell# show interfaces dampening summary
20 interfaces are configured with dampening. 3 interfaces are currently
suppressed.
Following interfaces are currently suppressed:
Gi 0/2
Gi 3/1
Gi 4/2
Dell#
```

Clearing Dampening Counters

To clear dampening counters and accumulated penalties, use the following command.

- Clear dampening counters.

  `clear dampening`

Example of the `clear dampening` Command

```
Dell# clear dampening interface Gi 0/1

Dell# show interfaces dampening GigabitEthernet0/0
InterfaceStateFlapsPenaltyHalf-LifeReuseSuppressMax-Sup
Gi 0/1Up00205001500300
```

Link Dampening Support for XML

View the output of the following `show` commands in XML by adding `| display xml` to the end of the command.

- `show interfaces dampening`
- `show interfaces dampening summary`
- `show interfaces interface x/y`

Configure MTU Size on an Interface

In Dell Networking OS, Maximum Transmission Unit (MTU) is defined as the entire Ethernet packet (Ethernet header + FCS + payload).

The link MTU is the frame size of a packet, and the IP MTU size is used for IP fragmentation. If the system determines that the IP packet must be fragmented as it leaves the interface, Dell Networking OS divides the packet into fragments no bigger than the size set in the `ip mtu` command.

NOTE: Because different networking vendors define MTU differently, check their documentation when planning MTU sizes across a network.

The following table lists the range for each transmission media.

Transmission Media	MTU Range (in bytes)
Ethernet	594-12000 = link MTU 576-9234 = IP MTU

Link Bundle Monitoring

Link bundle monitoring is supported only on the platform.

Monitoring linked LAG bundles allows traffic distribution amounts in a link to be monitored for unfair distribution at any given time. A threshold of 60% is defined as an acceptable amount of traffic on a member link. Links are monitored in 15-second intervals for three consecutive instances. Any deviation within that time sends Syslog and an alarm event generates. When the deviation clears, another Syslog sends and a clear alarm event generates.

The link bundle utilization is calculated as the total bandwidth of all links divided by the total bytes-per-second of all links. If you enable monitoring, the utilization calculation is performed when the utilization of the link-bundle (not a link within a bundle) exceeds 60%.

To enable and view link bundle monitoring, use the following commands.

- Enable link bundle monitoring.
  `ecmp-group`
- View all LAG link bundles being monitored.
  `show running-config ecmp-group`

Using Ethernet Pause Frames for Flow Control

Ethernet pause frames and threshold settings are supported on the S6000 platform.

Ethernet Pause Frames allow for a temporary stop in data transmission. A situation may arise where a sending device may transmit data faster than a destination device can accept it. The destination sends a PAUSE frame back to the source, stopping the sender's transmission for a period of time.

An Ethernet interface starts to send pause frames to a sending device when the transmission rate of ingress traffic exceeds the egress port speed. The interface stops sending pause frames when the ingress rate falls to less than or equal to egress port speed.

The globally assigned 48-bit Multicast address 01-80-C2-00-00-01 is used to send and receive pause frames. To allow full-duplex flow control, stations implementing the pause operation instruct the MAC to enable reception of frames with destination address equal to this multicast address.

The PAUSE frame is defined by IEEE 802.3x and uses MAC Control frames to carry the PAUSE commands. Ethernet pause frames are supported on full duplex only.

If a port is over-subscribed, Ethernet Pause Frame flow control does not ensure no-loss behavior.

Restriction: Ethernet Pause Frame flow control is not supported if PFC is enabled on an interface.

Control how the system responds to and generates 802.3x pause frames on Ethernet interfaces. The default is rx off tx off. INTERFACE mode. `flowcontrol rx [off | on] tx [off | on]`

Where:

`rx on`: Processes the received flow control frames on this port.

`rx off`: Ignores the received flow control frames on this port.

`tx on`: Sends control frames from this port to the connected device when a higher rate of traffic is received.

`tx off`: Flow control frames are not sent from this port to the connected device when a higher rate of traffic is received.

Changes in the flow-control values may not be reflected automatically in **show interface** output. To display the change, apply the new flow-control setting, perform a **shutdown** followed by a **no shutdown** on the interface, and then check re-display the **show interface** output for the port.

Threshold Settings

Threshold settings are supported on the S6000 platform.

When the transmission pause is set (`tx on`), you can set three thresholds to define the controls more closely. Ethernet pause frames flow control can be triggered when either the flow control buffer threshold or flow control packet pointer threshold is reached. The thresholds are:

- Number of flow-control packet pointers: from 1 to 2047 (default = **75**)
- Flow-control buffer threshold in KB: from 1 to 2013 (default = **49KB**)
- Flow-control discard threshold in KB: from 1-2013 (default= **75KB**)

The pause is started when *either* the packet pointer or the buffer threshold is met (whichever is met first). When the discard threshold is met, packets are dropped.

The pause ends when *both* the packet pointer and the buffer threshold fall below 50% of the threshold settings.

The discard threshold defines when the interface starts dropping the packet on the interface. This may be necessary when a connected device does not honor the flow control frame sent by the .

The discard threshold should be larger than the buffer threshold so that the buffer holds at least hold at least three packets.

Enabling Pause Frames

Enable Ethernet pause frames flow control on all ports on a chassis or a line card. If not, the system may exhibit unpredictable behavior.

NOTE: Changes in the flow-control values may not be reflected automatically in the `show interface` output. As a workaround, apply the new settings, execute `shut` then `no shut` on the interface, and then check the running-config of the port.

NOTE: If you disable `rx flow control`, Dell Networking recommends rebooting the system.

The flow control sender and receiver must be on the same port-pipe. Flow control is not supported across different port-pipes.

To enable pause frames, use the following command.

- Control how the system responds to and generates 802.3x pause frames on 1 and 10Gig line cards.

  INTERFACE mode

  ```
  flowcontrol rx [off | on] tx [off | on] [threshold {<1-2047> <1-2013>
  <1-2013>}]
  ```

  - `rx on`: enter the keywords `rx on` to process the received flow control frames on this port.
  - `rx off`: enter the keywords `rx off` to ignore the received flow control frames on this port.
  - `tx on`: enter the keywords `tx on` to send control frames from this port to the connected device when a higher rate of traffic is received.
  - `tx off`: enter the keywords `tx off` so that flow control frames are not sent from this port to the connected device when a higher rate of traffic is received.
  - `threshold`: when you configure `tx on`, you can set the threshold values for:
    - Number of flow-control packet pointers: the range is from 1 to 2047 (default = **75**).
    - Flow-control buffer threshold in KB: the range is from 1 to 2013 (default = **49KB**).
    - Flow-control discard threshold in KB: the range is from 1 to 2013 (default= **75KB**)

  Pause control is triggered when either the flow control buffer threshold or flow control packet pointer threshold is reached.

Configure the MTU Size on an Interface

If a packet includes a Layer 2 header, the difference in bytes between the link MTU and IP MTU must be enough to include the Layer 2 header.

For example, for VLAN packets, if the IP MTU is 1400, the Link MTU must be no less than 1422:

```
1400-byte IP MTU + 22-byte VLAN Tag = 1422-byte link MTU
```

The MTU range is from 592 to 12000, with a default of 1500. IP MTU automatically configures.

The following table lists the various Layer 2 overheads found in Dell Networking OS and the number of bytes.

Table 29. Layer 2 Overhead

Layer 2 Overhead	Difference Between Link MTU and IP MTU
Ethernet (untagged)	18 bytes
VLAN Tag	22 bytes
Untagged Packet with VLAN-Stack Header	22 bytes
Tagged Packet with VLAN-Stack Header	26 bytes

Link MTU and IP MTU considerations for port channels and VLANs are as follows.

Port Channels:

- All members must have the same link MTU value and the same IP MTU value.
- The port channel link MTU and IP MTU must be less than or equal to the link MTU and IP MTU values configured on the channel members.

For example, if the members have a link MTU of 2100 and an IP MTU 2000, the port channel's MTU values cannot be higher than 2100 for link MTU or 2000 bytes for IP MTU.

VLANs:

- All members of a VLAN must have the same IP MTU value.
- Members can have different Link MTU values. Tagged members must have a link MTU 4–bytes higher than untagged members to account for the packet tag.
- The VLAN link MTU and IP MTU must be less than or equal to the link MTU and IP MTU values configured on the VLAN members.

For example, the VLAN contains tagged members with Link MTU of 1522 and IP MTU of 1500 and untagged members with Link MTU of 1518 and IP MTU of 1500. The VLAN's Link MTU cannot be higher than 1518 bytes and its IP MTU cannot be higher than 1500 bytes.

Port-Pipes

A port pipe is a Dell Networking-specific term for the hardware path that packets follow through a system.

Port pipes travel through a collection of circuits (ASICs) built into line cards and RPMs on which various processing events for the packets occur. One or two port pipes process traffic for a given set of physical interfaces or a port-set.

Auto-Negotiation on Ethernet Interfaces

By default, auto-negotiation of speed and duplex mode is enabled on 10/100/1000 Base-T Ethernet interfaces. Only 10GE interfaces do not support auto-negotiation.

When using 10GE interfaces, verify that the settings on the connecting devices are set to no auto-negotiation.

NOTE: When you use a copper SFP2 module with catalog number GP-SFP2-1T in the S25P model, you can manually set its speed with the `speed` command. When the speed is set to 10Mbps or 100Mbps, you can execute the duplex command.

The local interface and the directly connected remote interface must have the same setting, and auto-negotiation is the easiest way to accomplish that, as long as the remote interface is capable of auto-negotiation.

NOTE: As a best practice, Dell Networking recommends keeping auto-negotiation enabled. Only disable auto-negotiation on switch ports that attach to devices not capable of supporting negotiation or where connectivity issues arise from interoperability issues.

For 10/100/1000 Ethernet interfaces, the `negotiation auto` command is tied to the `speed` command. Auto-negotiation is always enabled when the `speed` command is set to 1000 or auto.

Setting the Speed and Duplex Mode of Ethernet Interfaces

To discover whether the remote and local interface requires manual speed synchronization, and to manually synchronize them if necessary, use the following command sequence.

1. Determine the local interface status. Refer to the following example.

   EXEC Privilege mode

   `show interfaces [`*`interface`*` | linecard `*`slot-number`*`] status`
2. Determine the remote interface status.

   EXEC mode or EXEC Privilege mode

   [Use the command on the remote system that is equivalent to the first command.]
3. Access CONFIGURATION mode.

   EXEC Privilege mode

   `config`
4. Access the port.

   CONFIGURATION mode

   `interface `*`interface slot/port`*
5. Set the local port speed.

   INTERFACE mode

   `speed {10 | 100 | 1000 | auto}`
6. Optionally, set full- or half-duplex.

   INTERFACE mode

   `duplex {half | full}`
7. Disable auto-negotiation on the port.

   INTERFACE mode

   `no negotiation auto`

   If the speed was set to 1000, do not disable auto-negotiation.
8. Verify configuration changes.

   INTERFACE mode

   `show config`

Example of the `show interfaces status` Command to View Link Status

NOTE: The `show interfaces status` command displays link status, but not administrative status. For both link and administrative status, use the `show ip interface [interface | brief | linecard slot-number] [configuration]` command.

```
Dell#show interfaces status
Port Description Status Speed      Duplex Vlan
Gi 0/0           Up     1000  Mbit Auto   --
Gi 0/1           Down   Auto       Auto   1
Gi 0/2           Down   Auto       Auto   1
Gi 0/3           Down   Auto       Auto   --
Gi 0/4 Force10Port Up   1000 Mbit  Auto   30-130
Gi 0/5           Down   Auto       Auto   --
Gi 0/6           Down   Auto       Auto   --
Gi 0/7           Up     1000 Mbit  Auto   1502,1504,1506-1508,1602
Gi 0/8           Down   Auto       Auto   --
Gi 0/9           Down   Auto       Auto   --
Gi 0/10          Down   Auto       Auto   --
Gi 0/11          Down   Auto       Auto   --
Gi 0/12          Down   Auto       Auto   --
[output omitted]
```

In the previous example, several ports display "Auto" in the Speed field, including port 0/1. In the following example, the speed of port 0/1 is set to 100Mb and then its auto-negotiation is disabled.

```
Dell#configure
Dell(config)#interface tengig 0/1
Dell(conf-if-te-0/1)#speed 100
Dell(conf-if-te-0/1)#duplex full
Dell(conf-if-te-0/1)#no negotiation auto
Dell(conf-if-te-0/1)#show config
!
interface GigabitEthernet 0/1
no ip address
speed 100
duplex full
no shutdown
```

Set Auto-Negotiation Options

The `negotiation auto` command provides a mode option for configuring an individual port to forced master/ forced slave once auto-negotiation is enabled.

CAUTION: Ensure that only one end of the node is configured as forced-master and the other is configured as forced-slave. If both are configured the same (that is, both as forced-master or both as forced-slave), the `show interface` command flaps between an auto-neg-error and forced-master/slave states.

Example of the `negotiation auto` Command

```
Dell(conf)# int tengig 0/0
Dell(conf-if-te-0/1)#neg auto
Dell(conf-if-te-0/1)# ?

end              Exit from configuration mode
exit             Exit from autoneg configuration mode
mode        Specify autoneg mode
no            Negate a command or set its defaults
show             Show autoneg configuration information
Dell(conf-if-te-0/1)#mode ?
forced-master Force port to master mode
forced-slave     Force port to slave mode
Dell(conf-if-te-0/1)#
```

For details about the `speed`, `duplex`, and `negotiation auto` commands, refer to the *Interfaces* chapter of the *Dell Networking OS Command Reference Guide*.

Adjusting the Keepalive Timer

To change the time interval between keepalive messages on the interfaces, use the `keepalive` command. The interface sends keepalive messages to itself to test network connectivity on the interface. To change the default time interval between keepalive messages, use the following command.

- Change the default interval between keepalive messages.

  INTERFACE mode

  ```
  keepalive [seconds]
  ```
- View the new setting.

  INTERFACE mode

  ```
  show config
  ```

View Advanced Interface Information

The following options have been implemented for the `show [ip | running-config] interfaces` commands for (only) stack-unit interfaces.

When you use the `configured` keyword, only interfaces that have non-default configurations are displayed. Dummy stack-unit interfaces (created with the `stack-unit` command) are treated like any other physical interface.

Examples of the `show` Commands

The following example lists the possible `show` commands that have the configured keyword available:

```
Dell#show interfaces configured
Dell#show interfaces stack-unit 0 configured
Dell#show interfaces tengigabitEthernet 0 configured
Dell#show ip interface configured
Dell#show ip interface stack-unit 1 configured
Dell#show ip interface tengigabitEthernet 1 configured
Dell#show ip interface br configured
Dell#show ip interface br stack-unit 1 configured
Dell#show ip interface br tengigabitEthernet 1 configured
Dell#show running-config interfaces configured
Dell#show running-config interface tengigabitEthernet 1 configured
```

In EXEC mode, the `show interfaces switchport` command displays only interfaces in Layer 2 mode and their relevant configuration information. The `show interfaces switchport` command displays the interface, whether it supports IEEE 802.1Q tagging or not, and the VLANs to which the interface belongs.

```
Dell#show interfaces switchport
Name: TenGigabitEthernet 13/0
802.1QTagged: True
Vlan membership:
Vlan 2

Name: TenGigabitEthernet 13/1
802.1QTagged: True
Vlan membership:
Vlan 2

Name: TenGigabitEthernet 13/2
802.1QTagged: True
Vlan membership:
Vlan 2

Name: TenGigabitEthernet 13/3
802.1QTagged: True
Vlan membership:
Vlan 2

--More--
```

Configuring the Interface Sampling Size

Although you can enter any value between 30 and 299 seconds (the default), software polling is done once every 15 seconds. So, for example, if you enter “19”, you actually get a sample of the past 15 seconds.
All LAG members inherit the rate interval configuration from the LAG.

The following example shows how to configure rate interval when changing the default value.

To configure the number of seconds of traffic statistics to display in the show interfaces output, use the following command.

- Configure the number of seconds of traffic statistics to display in the show interfaces output.

  INTERFACE mode

  `rate-interval`

Example of the `rate-interval` Command

The bold lines shows the default value of 299 seconds, the change-rate interval of 100, and the new rate interval set to 100.

```
Dell#show interfaces
TenGigabitEthernet 10/0 is down, line protocol is down
Hardware is Force10Eth, address is 00:01:e8:01:9e:d9
Internet address is not set
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 10000 Mbit
ARP type: ARPA, ARP Timeout 04:00:00
Last clearing of "show interface" counters 1d23h44m
Queueing strategy: fifo
  0 packets input, 0 bytes
  Input 0 IP Packets, 0 Vlans 0 MPLS
  0 64-byte pkts, 0 over 64-byte pkts, 0 over 127-byte pkts
  0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
  Received 0 input symbol errors, 0 runts, 0 giants, 0 throttles
  0 CRC, 0 IP Checksum, 0 overrun, 0 discarded
  0 packets output, 0 bytes, 0 underruns
  Output 0 Multicasts, 0 Broadcasts, 0 Unicasts
  0 IP Packets, 0 Vlans, 0 MPLS
  0 throttles, 0 discarded
Rate info (interval 299 seconds):
  Input 00.00 Mbits/sec,  0 packets/sec, 0.00% of line-rate
  Output 00.00 Mbits/sec, 0 packets/sec, 0.00% of line-rate
Time since last interface status change: 1d23h40m

Dell(conf)#interface tengigabitethernet 10/0
Dell(conf-if-te-10/0)#rate-interval 100

Dell#show interfaces
TenGigabitEthernet 10/0 is down, line protocol is down
Hardware is Force10Eth, address is 00:01:e8:01:9e:d9
Internet address is not set
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 10000 Mbit
ARP type: ARPA, ARP Timeout 04:00:00
Last clearing of "show interface" counters 1d23h45m
Queueing strategy: fifo
  0 packets input, 0 bytes
  Input 0 IP Packets, 0 Vlans 0 MPLS
  0 64-byte pkts, 0 over 64-byte pkts, 0 over 127-byte pkts
  0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
Received 0 input symbol errors, 0 runts, 0 giants, 0 throttles
  0 CRC, 0 IP Checksum, 0 overrun, 0 discarded
  0 packets output, 0 bytes, 0 underruns
  Output 0 Multicasts, 0 Broadcasts, 0 Unicasts
  0 IP Packets, 0 Vlans, 0 MPLS
  0 throttles, 0 discarded
Rate info (interval 100 seconds):
  Input 00.00 Mbits/sec,  0 packets/sec, 0.00% of line-rate
  Output 00.00 Mbits/sec, 0 packets/sec, 0.00% of line-rate
Time since last interface status change: 1d23h42m
```

Dynamic Counters

By default, counting is enabled for IPFLOW, IPACL, L2ACL, L2FIB.

For the remaining applications, Dell Networking OS automatically turns on counting when you enable the application, and is turned off when you disable the application.

NOTE: If you enable more than four counter-dependent applications on a port pipe, there is an impact on line rate performance.

The following counter-dependent applications are supported by Dell Networking OS:

- Egress VLAN
- Ingress VLAN
- Next Hop 2
- Next Hop 1
- Egress ACLs
- ILM
- IP FLOW
- IP ACL
- IP FIB
- L2 ACL
- L2 FIB

Clearing Interface Counters

The counters in the `show interfaces` command are reset by the `clear counters` command. This command does not clear the counters any SNMP program captures.
To clear the counters, use the following the command.

- Clear the counters used in the `show interface` commands for all VRRP groups, VLANs, and physical interfaces or selected ones. Without an interface specified, the command clears all interface counters.

  EXEC Privilege mode

  ```
  clear counters [interface] [vrrp [vrid] | learning-limit]
  ```

  (OPTIONAL) Enter the following interface keywords and slot/port or number information:

  - For a 1-Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
  - For a Loopback interface, enter the keyword `loopback` then a number from 0 to 16383.

  - For a Port Channel interface, enter the keywords `port-channel` then a number.
  - For the management interface on the RPM, enter the keyword `ManagementEthernet` then the slot/port information. The slot range is from 0 to 1. The port range is 0.
  - For a SONET interface, enter the keyword `sonet` then the slot/ port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
  - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.
  - For a VLAN, enter the keyword `vlan` then a number.
  - (OPTIONAL) To clear statistics for all VRRP groups configured, enter the keyword `vrrp`. Enter a number from 1 to 255 as the *`vrid`*.
  - (OPTIONAL) To clear unknown source address (SA) drop counters when you configure the MAC learning limit on the interface, enter the keywords `learning-limit`.

Example of the `clear counters` Command

When you enter this command, confirm that you want Dell Networking OS to clear the interface counters for that interface.

```
Dell#clear counters gi 0/0
Clear counters on GigabitEthernet 0/0 [confirm]
Dell#
```

Enhanced Validation of Interface Ranges

This functionality is supported on the S6000 platform.

You can avoid specifying spaces between the range of interfaces, separated by commas, that you configure by using the `interface range` command. For example, if you enter a list of interface ranges, such as `interface range fo 2/0-1,te 10/0,gi 3/0,fa 0/0`, this configuration is considered valid. The comma-separated list is not required to be separated by spaces in between the ranges. You can associate multicast MAC or hardware addresses to an interface range and VLANs by using the `mac-address-table static` *`multicast-mac-address`* `vlan` *`vlan-id`* `output-range` *`interface`* command.

Compressing Configuration Files

The functionality to optimize and reduce the sizes of the configuration files is supported on the S6000 platform.

You can compress the running configuration by grouping all the VLANs and the physical interfaces with the same property. Support to store the operating configuration to the startup config in the compressed mode and to perform an image downgrade without any configuration loss are provided.

You can create groups of VLANs using the `interface group` command. This command will create non-existant VLANs specified in a range. On successful command execution, the CLI switches to the interface group context. The configuration commands inside the group context will be the similar to that of the existing range command.

Two existing exec mode CLIs are enhanced to display and store the running configuration in the compressed mode.

```
show running-config compressed and write memory compressed
```

The compressed configuration will group all the similar looking configuration thereby reducing the size of the configuration. For this release, the compression will be done only for interface related configuration (VLAN & physical interfaces)

The following table describes how the standard and the compressed configuration differ:

int vlan 2 no ip address no shut	int vlan 3 tagged te 0/0 no ip address shut	int vlan 4 tagged te 0/0 no ip address shut	int vlan 5 tagged te 0/0 no ip address shut	int vlan 100 no ip address no shut	int vlan 1000 ip address 1.1.1.1/16 no shut
int te 0/ 0 no ip address switchport shut	int te 0/2 no ip address shut	int te 0/3 no ip address shut	int te 0/4 no ip address shut	int te 0/10 no ip address shut	int te 0/34 ip address 2.1.1.1/16 shut

Dell# show running-config	Dell# show running-config compressed
<snip> ! interface TenGigabitEthernet 0/0 no ip address switchport shutdown ! interface TenGigabitEthernet 0/2 no ip address shutdown ! interface TenGigabitEthernet 0/3 no ip address shutdown ! interface TenGigabitEthernet 0/4 no ip address	<snip> ! interface TenGigabitEthernet 0/0 no ip address switchport shutdown ! Interface group TenGigabitEthernet 0/2 – 4 , TenGigabitEthernet 0/10 no ip address shutdown ! interface TenGigabitEthernet 0/34 ip address 2.1.1.1/16 shutdown ! interface group Vlan 2 , Vlan 100 no ip address

```
shutdown
!
interface TenGigabitEthernet 0/10
no ip address
shutdown
!
interface TenGigabitEthernet 0/34
ip address 2.1.1.1/16
shutdown
!
interface Vlan 2
no ip address
no shutdown
!
interface Vlan 3
tagged te 0/0
no ip address
shutdown
!
interface Vlan 4
tagged te 0/0
no ip address
shutdown
!
interface Vlan 5
tagged te 0/0
no ip address
shutdown
!
interface Vlan 100
no shutdown
!
interface group Vlan 3 – 5
tagged te 0/0
no ip address
shutdown
!
interface Vlan 1000
ip address 1.1.1.1/16
no shutdown
!
<snip>
Compressed config size – 27 lines.
```

no ip address

no shutdown

!

interface Vlan 1000

ip address 1.1.1.1/16

no shutdown

Uncompressed config size – 52 lines

write memory compressed

The write memory compressed CLI will write the operating configuration to the startup-config file in the compressed mode. In stacking scenario, it will also take care of syncing it to all the standby and member units.

The following is the sample output:

```
Dell#write memory compressed
!
Jul 30 08:50:26: %STKUNIT0-M:CP %FILEMGR-5-FILESAVED: Copied running-config to
startup-config in flash by default
```

copy compressed-config

Copy one file, after optimizing and reducing the size of the configuration file, to another location. Dell Networking OS supports IPv4 and IPv6 addressing for FTP, TFTP, and SCP (in the hostip field).

20

IPv4 Routing

IPv4 routing is supported on the S6000 platform.

The Dell Networking Operating System (OS) supports various IP addressing features. This chapter describes the basics of domain name service (DNS), address resolution protocol (ARP), and routing principles and their implementation in the Dell Networking OS.

IP Feature	Default
DNS	Disabled
Directed Broadcast	Disabled
Proxy ARP	Enabled
ICMP Unreachable	Disabled
ICMP Redirect	Disabled

IP Addresses

Dell Networking OS supports IP version 4, as described in RFC 791. Dell Networking OS also supports classful routing and variable length subnet masks (VLSM).

With VLSM, you can configure one network with different masks. Supernetting, which increases the number of subnets, is also supported. To subnet, you add a mask to the IP address to separate the network and host portions of the IP address.

At its most basic level, an IP address is 32-bits composed of network and host portions and represented in dotted decimal format. For example, 00001010110101100101011110000011 is represented as 10.214.87.131.

For more information about IP addressing, refer to RFC 791, Internet Protocol.

Implementation Information

In Dell Networking OS, you can configure any IP address as a static route except IP addresses already assigned to interfaces.

NOTE: Dell Networking OS supports 31-bit subnet masks (/31, or 255.255.255.254) as defined by RFC 3021. This feature allows you to save two more IP addresses on point-to-point links than 30-bit masks. Dell Networking OS supports RFC 3021 with ARP.

Configuration Tasks for IP Addresses

The following describes the tasks associated with IP address configuration.

Configuration tasks for IP addresses includes:

- Assigning IP Addresses to an Interface (mandatory)
- Configuring Static Routes (optional)
- Configure Static Routes for the Management Interface (optional)

For a complete listing of all commands related to IP addressing, refer to the *Dell Networking OS Command Line Interface Reference Guide*.

Assigning IP Addresses to an Interface

Assign primary and secondary IP addresses to physical or logical (for example, [virtual local area network [VLAN] or port channel) interfaces to enable IP communication between the system and hosts connected to that interface.
In Dell Networking OS, you can assign one primary address and up to 255 secondary IP addresses to each interface.

1. Enter the keyword `interface` then the type of interface and slot/port information.

   CONFIGURATION mode

   ```
   interface interface
   ```

   - For a 1-Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
   - For a Loopback interface, enter the keyword `loopback` then a number from 0 to 16383.
   - For the Management interface on the RPM, enter the keyword `ManagementEthernet` then the slot/port information. The slot range is from 0 to 1. The port range is 0.
   - For a port channel interface, enter the keywords `port-channel` then a number.
   - For a SONET interface, enter the keyword `sonet` then the slot/port information.
   - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
   - For a VLAN interface, enter the keyword `vlan` then a number from 1 to 4094.
   - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.
2. Enable the interface.

   INTERFACE mode

   ```
   no shutdown
   ```
3. Configure a primary IP address and mask on the interface.

   INTERFACE mode

   ```
   ip address ip-address mask [secondary]
   ```

   - *`ip-address mask`*: the IP address must be in dotted decimal format (A.B.C.D). The mask must be in slash prefix-length format (/24).
   - `secondary`: add the keyword `secondary` if the IP address is the interface’s backup IP address. You can configure up to eight secondary IP addresses.

Example the `show config` Command

To view the configuration, use the `show config` command in INTERFACE mode or use the `show ip interface` command in EXEC privilege mode, as shown in the second example.

```
Dell(conf-if)#show conf
!
```

```
interface GigabitEthernet 0/0
  ip address 10.11.1.1/24
  no shutdown
!
Dell(conf-if)#
```

```
Dell(conf-if)#show conf
!
interface GigabitEthernet 0/0
ip address 10.11.1.1/24
no shutdown
!

Dell(conf-if)#
```

Configuring Static Routes

A static route is an IP address that you manually configure and that the routing protocol does not learn, such as open shortest path first (OSPF). Often, static routes are used as backup routes in case other dynamically learned routes are unreachable.
You can enter as many static IP addresses as necessary.

To configure a static route, use the following command.

- Configure a static IP address.

  CONFIGURATION mode

  `ip route [vrf` *`vrf-name`*`]` *`ip-address mask`* `{`*`ip-address`* `|` *`interface`* `[`*`ip-address`*`]} [`*`distance`*`] [permanent] [tag` *`tag-value`*`] [vrf` *`vrf-name`*`]`

  Use the following required and optional parameters:

  - `vrf` *`vrf-name`* : use the VRF option after the `ip route` keyword to configure a static route on that particular VRF, use the VRF option after the next hop to specify which VRF the next hop belongs to. This will be used in route leaking cases.

    NOTE: For more information on route leaking, see the Route Leaking Between VRFs section.

  - *`ip-address`*: enter an address in dotted decimal format (A.B.C.D).
  - *`mask`*: enter a mask in slash prefix-length format (/X).
  - *`interface`*: enter an interface type then the slot/port information.
  - *`distance`*: the range is from 1 to 255. (optional)
  - `permanent`: keep the static route in the routing table (if you use the `interface` option) even if you disable the interface with the route. (optional)
  - `tag` *`tag-value`*: the range is from 1 to 4294967295. (optional)

Example of the `show ip route static` Command

To view the configured routes, use the `show ip route static` command.

```
Dell#show ip route static
  Destination   Gateway                Dist/Metric Last Change
-----------     -------                ----------- -----------
S 2.1.2.0/24    Direct, Nu 0              0/0      00:02:30
S 6.1.2.0/24    via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.2/32    via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.3/32    via 6.1.20.2, Te 5/0      1/0      00:02:30
```

```
S 6.1.2.4/32   via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.5/32   via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.6/32   via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.7/32   via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.8/32   via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.9/32   via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.10/32  via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.11/32  via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.12/32  via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.13/32  via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.14/32  via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.15/32  via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.16/32  via 6.1.20.2, Te 5/0      1/0      00:02:30
S 6.1.2.17/32  via 6.1.20.2, Te 5/0      1/0      00:02:30
S 11.1.1.0/24  Direct, Nu 0              0/0      00:02:30
Direct, Lo 0
--More--
```

Dell Networking OS installs a next hop that is on the directly connected subnet of current IP address on the interface (for example, if `interface gig 0/0` is on 172.31.5.0 subnet, Dell Networking OS installs the static route).

Dell Networking OS also installs a next hop that is not on the directly connected subnet but which recursively resolves to a next hop on the interface's configured subnet. For example, if `gig 0/0` has ip address on subnet 2.2.2.0 and if 172.31.5.43 recursively resolves to 2.2.2.0, Dell Networking OS installs the static route.

- When the interface goes down, Dell Networking OS withdraws the route.
- When the interface comes up, Dell Networking OS re-installs the route.
- When the recursive resolution is "broken," Dell Networking OS withdraws the route.
- When the recursive resolution is satisfied, Dell Networking OS re-installs the route.

Configure Static Routes for the Management Interface

When an IP address that a protocol uses and a static management route exists for the same prefix, the protocol route takes precedence over the static management route.
To configure a static route for the management port, use the following command.

- Assign a static route to point to the management interface or forwarding router.

  CONFIGURATION mode

  ```
  management route ip-address mask {forwarding-router-address |
  ManagementEthernet slot/port}
  ```

Example of the `show ip route static` Command

To view the configured static routes for the management port, use the `show ip management-route` command in EXEC privilege mode.

```
Dell#show ip route static
  Destination   Gateway             Dist/Metric Last Change
-----------     -------             ----------- -----------
S 2.1.2.0/24    Direct, Nu 0           0/0      00:02:30
S 6.1.2.0/24    via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.2/32    via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.3/32    via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.4/32    via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.5/32    via 6.1.20.2, Te 5/0   1/0      00:02:30
```

```
S 6.1.2.6/32    via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.7/32    via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.8/32    via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.9/32    via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.10/32   via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.11/32   via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.12/32   via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.13/32   via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.14/32   via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.15/32   via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.16/32   via 6.1.20.2, Te 5/0   1/0      00:02:30
S 6.1.2.17/32   via 6.1.20.2, Te 5/0   1/0      00:02:30
S 11.1.1.0/24   Direct, Nu 0           0/0      00:02:30
                Direct, Lo 0
--More--
```

Using the Configured Source IP Address in ICMP Messages

This feature is supported on the S6000 platform.

ICMP error or unreachable messages are now sent with the configured IP address of the source interface instead of the front-end port IP address as the source IP address. Enable the generation of ICMP unreachable messages through the `ip unreachable` command in Interface mode. When a ping or traceroute packet from an endpoint or a device arrives at the null 0 interface configured with a static route, it is discarded. In such cases, you can configure Internet Control Message Protocol (ICMP) unreachable messages to be sent to the transmitting device.

Configuring the ICMP Source Interface

You can enable the ICMP error and unreachable messages to contain the configured IP address of the source device instead of the previous hop's IP address. This configuration helps identify the devices along the path because the DNS server maps the loopback IP address to the host name, and does not translate the IP address of every interface of the switch to the host name.

Configure the source to send the configured source interface IP address instead of using its front-end IP address in the ICMP unreachable messages and in the `traceroute` command output. Use the `ip icmp source-interface` *`interface`* or the `ipv6 icmp source-interface` *`interface`* commands in Configuration mode to enable the ICMP error messages to be sent with the source interface IP address. This functionality is supported on loopback, VLAN, port channel, and physical interfaces for IPv4 and IPv6 messages. feature is not supported on tunnel interfaces. ICMP error relay, PATH MTU transmission, and fragmented packets are not supported for tunnel interfaces. The traceroute utilities for IPv4 and IPv6 list the IP addresses of the devices in the hops of the path for which ICMP source interface is configured.

Configuring the Duration to Establish a TCP Connection

This functionality is supported on the S6000 platform.

You can configure the amount of time for which the device must wait before it attempts to establish a TCP connection. Using this capability, you can limit the wait times for TCP connection requests. Upon responding to the initial SYN packet that requests a connection to the router for a specific service (such as SSH or BGP) with a SYN ACK, the router waits for a period of time for the ACK packet to be sent from the requesting host that will establish the TCP connection.

You can set this duration or interval for which the TCP connection waits to be established to a significantly high value to prevent the device from moving into an out-of-service condition or becoming unresponsive during a SYN flood attack that occurs on the device. You can set the wait time to be 10 seconds or lower. If the device does not contain any BGP connections with the BGP neighbors across WAN links, you must set this interval to a higher value, depending on the complexity of your network and the configuration attributes.

To configure the duration for which the device waits for the ACK packet to be sent from the requesting host to establish the TCP connection, perform the following steps:

1. Define the wait duration in seconds for the TCP connection to be established.
   CONFIGURATION mode

   ```
   Dell(conf)#ip tcp reduced-syn-ack-wait <9-75>
   ```

   You can use the `no ip tcp reduced-syn-ack-wait` command to restore the default behavior, which causes the wait period to be set as 8 seconds.
2. View the interval that you configured for the device to wait before the TCP connection is attempted to be established.
   EXEC mode

   ```
   Dell>show ip tcp reduced-syn-ack-wait
   ```

Enabling Directed Broadcast

By default, Dell Networking OS drops directed broadcast packets destined for an interface. This default setting provides some protection against denial of service (DoS) attacks.
To enable Dell Networking OS to receive directed broadcasts, use the following command.

- Enable directed broadcast.
  INTERFACE mode

  ```
  ip directed-broadcast
  ```

To view the configuration, use the `show config` command in INTERFACE mode.

Resolution of Host Names

Domain name service (DNS) maps host names to IP addresses. This feature simplifies such commands as Telnet and FTP by allowing you to enter a name instead of an IP address.

Dynamic resolution of host names is disabled by default. Unless you enable the feature, the system resolves only host names entered into the host table with the `ip host` command.

The following sections describe DNS and the resolution of host names.

- Enabling Dynamic Resolution of Host Names
- Specifying the Local System Domain and a List of Domains
- Configuring DNS with Traceroute

In the release 9.4.(0.0), the Name server, Domain name, and Domain list are VRF specific. The maximum number of Name servers and Domain lists per VRF is six.

Enabling Dynamic Resolution of Host Names

By default, dynamic resolution of host names (DNS) is disabled.
To enable DNS, use the following commands.

- Enable dynamic resolution of host names.

  CONFIGURATION mode

  ```
  ip domain-lookup
  ```
- Specify up to six name servers.

  CONFIGURATION mode

  ```
  ip name-server ip-address [ip-address2 ... ip-address6]
  ```

  The order you entered the servers determines the order of their use.

Example of the `show hosts` Command

To view current bindings, use the `show hosts` command.

```
Dell>show host
Default domain is force10networks.com
Name/address lookup uses domain service
Name servers are not set
Host      Flags TTL     Type Address
-------- ----- ----     ---- -------
ks        (perm, OK) -  IP   2.2.2.2
patch1    (perm, OK) -  IP   192.68.69.2
tomm-3    (perm, OK) -  IP   192.68.99.2
gxr       (perm, OK) -  IP   192.71.18.2
f00-3     (perm, OK) -  IP   192.71.23.1
Dell>
```

To view the current configuration, use the `show running-config resolve` command.

Specifying the Local System Domain and a List of Domains

If you enter a partial domain, Dell Networking OS can search different domains to finish or fully qualify that partial domain.
A fully qualified domain name (FQDN) is any name that is terminated with a period/dot. Dell Networking OS searches the host table first to resolve the partial domain. The host table contains both statically configured and dynamically learnt host and IP addresses. If Dell Networking OS cannot resolve the domain, it tries the domain name assigned to the local system. If that does not resolve the partial domain, Dell Networking OS searches the list of domains configured.

To configure a domain name or a list of domain names, use the following commands.

- Enter up to 63 characters to configure one domain name.

  CONFIGURATION mode

  ```
  ip domain-name name
  ```
- Enter up to 63 characters to configure names to complete unqualified host names.

  CONFIGURATION mode

```
ip domain-list name
```

Configure this command up to six times to specify a list of possible domain names. Dell Networking OS searches the domain names in the order they were configured until a match is found or the list is exhausted.

Configuring DNS with Traceroute

To configure your switch to perform DNS with traceroute, use the following commands.

- Enable dynamic resolution of host names.

  CONFIGURATION mode

  ```
  ip domain-lookup
  ```
- Specify up to six name servers.

  CONFIGURATION mode

  ```
  ip name-server ip-address [ip-address2 ... ip-address6]
  ```

  The order you entered the servers determines the order of their use.
- When you enter the `traceroute` command without specifying an IP address (`Extended Traceroute`), you are prompted for a target and source IP address, timeout in seconds (default is **5**), a probe count (default is **3**), minimum TTL (default is **1**), maximum TTL (default is **30**), and port number (default is **33434**).

  CONFIGURATION mode

  ```
  traceroute [host | ip-address]
  ```

  To keep the default setting for these parameters, press the ENTER key.

Example of the `traceroute` Command

The following text is example output of DNS using the `traceroute` command.

```
Dell#traceroute www.force10networks.com

Translating "www.force10networks.com"...domain server (10.11.0.1) [OK]
Type Ctrl-C to abort.

----------------------------------------------------------------------
Tracing the route to www.force10networks.com (10.11.84.18), 30 hops max, 40
byte packets
----------------------------------------------------------------------

TTL Hostname        Probe1     Probe2     Probe3
1   10.11.199.190  001.000 ms 001.000 ms 002.000 ms
2   gwegress-sjc-02.force10networks.com (10.11.30.126) 005.000 ms 001.000 ms
001.000 ms
3   fw-sjc-01.force10networks.com (10.11.127.254) 000.000 ms 000.000 ms 000.000
ms
4   www.dell.com (10.11.84.18) 000.000 ms 000.000 ms 000.000 ms
Dell#
```

ARP

Dell Networking OS uses two forms of address resolution: address resolution protocol (ARP) and Proxy ARP.

ARP runs over Ethernet and enables endstations to learn the MAC addresses of neighbors on an IP network. Over time, Dell Networking OS creates a forwarding table mapping the MAC addresses to their corresponding IP address. This table is called the ARP Cache and dynamically learned addresses are removed after a defined period of time.

For more information about ARP, refer to RFC 826, *An Ethernet Address Resolution Protocol*.

In Dell Networking OS, Proxy ARP enables hosts with knowledge of the network to accept and forward packets from hosts that contain no knowledge of the network. Proxy ARP makes it possible for hosts to be ignorant of the network, including subnetting.

For more information about Proxy ARP, refer to RFC 925, *Multi-LAN Address Resolution*, and RFC 1027, *Using ARP to Implement Transparent Subnet Gateways*.

Configuration Tasks for ARP

For a complete listing of all ARP-related commands, refer to the *Dell Networking OS Command Line Reference Guide*.

Configuration tasks for ARP include:

- Configuring Static ARP Entries (optional)
- Enabling Proxy ARP (optional)
- Clearing ARP Cache (optional)
- ARP Learning via Gratuitous ARP
- ARP Learning via ARP Request
- Configuring ARP Retries

Configuring Static ARP Entries

ARP dynamically maps the MAC and IP addresses, and while most network host support dynamic mapping, you can configure an ARP entry (called a static ARP) for the ARP cache.
To configure a static ARP entry, use the following command.

- Configure an IP address and MAC address mapping for an interface.

  CONFIGURATION mode

  ```
  arp vrf vrf-name ip-address mac-address interface
  ```

  - `vrf vrf-name`: use the VRF option to configure a static ARP on that particular VRF.
  - `ip-address`: IP address in dotted decimal format (A.B.C.D).
  - `mac-address`: MAC address in nnnn.nnnn.nnnn format.
  - `interface`: enter the interface type slot/port information.

Example of the `show arp` Command

These entries do not age and can only be removed manually. To remove a static ARP entry, use the `no arp ip-address` command.

To view the static entries in the ARP cache, use the `show arp static` command in EXEC privilege mode.

```
Dell#show arp

Protocol  Address   Age(min) Hardware Address   Interface VLAN CPU
--------------------------------------------------------------------------------
Internet  10.1.2.4  17      08:00:20:b7:bd:32  Ma 1/0     -   CP
Dell#
```

Enabling Proxy ARP

By default, Proxy ARP is enabled. To disable Proxy ARP, use the `no proxy-arp` command in the interface mode.
To re-enable Proxy ARP, use the following command.

- Re-enable Proxy ARP.

  INTERFACE mode

  ```
  ip proxy-arp
  ```

To view if Proxy ARP is enabled on the interface, use the `show config` command in INTERFACE mode. If it is not listed in the `show config` command output, it is enabled. Only non-default information is displayed in the `show config` command output.

Clearing ARP Cache

To clear the ARP cache of dynamically learnt ARP information, use the following command.

- Clear the ARP caches for all interfaces or for a specific interface by entering the following information.

  EXEC privilege

  ```
  clear arp-cache [interface | ip ip-address] [no-refresh]
  ```

  - `ip` *`ip-address`* (OPTIONAL): enter the keyword `ip` then the IP address of the ARP entry you wish to clear.
  - `no-refresh` (OPTIONAL): enter the keywords `no-refresh` to delete the ARP entry from CAM. Or to specify which dynamic ARP entries you want to delete, use this option with *`interface`* or *`ip ip-address`*.
  - For a 1-Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
  - For a port channel interface, enter the keywords `port-channel` then a number.
  - For a SONET interface, enter the keyword `sonet` then the slot/port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
  - For a VLAN interface, enter the keyword `vlan` then a number between 1 and 4094.

  - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.

NOTE: Transit traffic may not be forwarded during the period when deleted ARP entries are resolved again and re-installed in CAM. Use this option with extreme caution.

ARP Learning via Gratuitous ARP

Gratuitous ARP can mean an ARP request or reply.

In the context of ARP learning via gratuitous ARP on Dell Networking OS, the gratuitous ARP is a request. A gratuitous ARP request is an ARP request that is not needed according to the ARP specification, but one that hosts may send to:

- detect IP address conflicts
- inform switches of their presence on a port so that packets can be forwarded
- update the ARP table of other nodes on the network in case of an address change

In the request, the host uses its own IP address in the Sender Protocol Address and Target Protocol Address fields.

In Dell Networking OS versions prior to 8.3.1.0, if a gratuitous ARP is received some time after an ARP request is sent, only RP2 installs the ARP information. For example:

1. At time t=0 Dell Networking OS sends an ARP request for IP *A.B.C.D*
2. At time t=1 Dell Networking OS receives an ARP request for IP *A.B.C.D*
3. At time t=2 Dell Networking OS installs an ARP entry for *A.B.C.D* only on RP2.

Beginning with Dell Networking OS version 8.3.1.0, when a gratuitous ARP is received, Dell Networking OS installs an ARP entry on all three CPUs.

Enabling ARP Learning via Gratuitous ARP

To enable ARP learning via gratuitous ARP, use the following command.

- Enable ARP learning via gratuitous ARP.
  CONFIGURATION mode

  ```
  arp learn-enable
  ```

ARP Learning via ARP Request

In Dell Networking OS versions prior to 8.3.1.0, Dell Networking OS learns via ARP requests only if the target IP specified in the packet matches the IP address of the receiving router interface. This is the case when a host is attempting to resolve the gateway address.

If the target IP does not match the incoming interface, the packet is dropped. If there is an existing entry for the requesting host, it is updated.

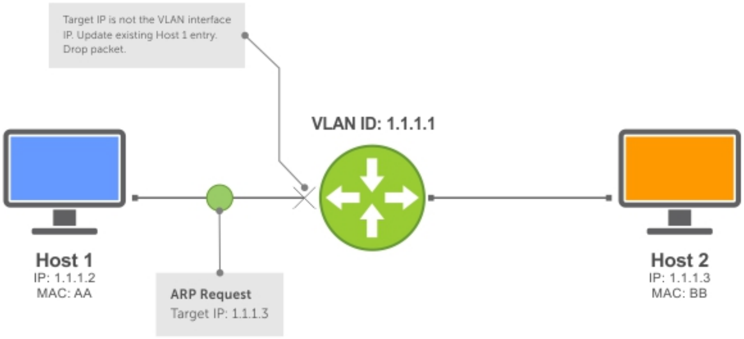

Target IP is not the VLAN interface IP. Update existing Host 1 entry. Drop packet.
VLAN ID: 1.1.1.1
Host 1
IP: 1.1.1.2
MAC: AA
ARP Request
Target IP: 1.1.1.3
Host 2
IP: 1.1.1.3
MAC: BB

Figure 42. ARP Learning via ARP Request

Beginning with Dell Networking OS version 8.3.1.0, when you enable ARP learning via gratuitous ARP, the system installs a new ARP entry, or updates an existing entry for all received ARP requests.

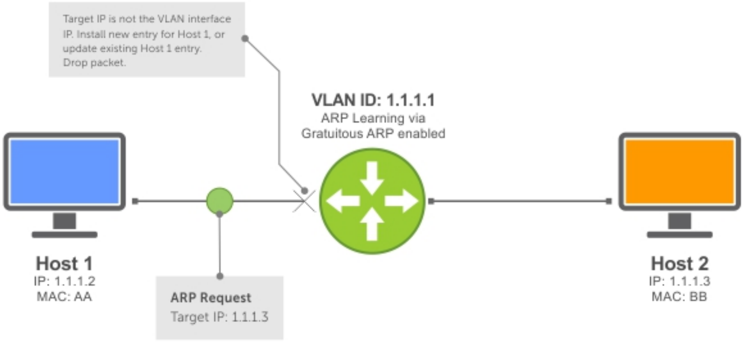

Target IP is not the VLAN interface IP. Install new entry for Host 1, or update existing Host 1 entry. Drop packet.
VLAN ID: 1.1.1.1
ARP Learning via Gratuitous ARP enabled
Host 1
IP: 1.1.1.2
MAC: AA
ARP Request
Target IP: 1.1.1.3
Host 2
IP: 1.1.1.3
MAC: BB

Figure 43. ARP Learning via ARP Request with ARP Learning via Gratuitous ARP Enabled

Whether you enable or disable ARP learning via gratuitous ARP, the system does not look up the target IP. It only updates the ARP entry for the Layer 3 interface with the source IP of the request.

Configuring ARP Retries

In Dell Networking OS versions prior to 8.3.1.0, the number of ARP retries is set to five and is not configurable. After five retries, Dell Networking OS backs off for 20 seconds before it sends a new request. Beginning with Dell Networking OS version 8.3.1.0, the number of ARP retries is configurable. The default backoff interval remains at 20 seconds.

To set and display ARP retries, use the following commands.

- Set the number of ARP retries.

  CONFIGURATION mode

  ```
  arp retries number
  ```

  The default is **5**.

The range is from 1 to 20.

- Set the exponential timer for resending unresolved ARPs.

  CONFIGURATION mode

  ```
  arp backoff-time
  ```

  The default is **30**.

  The range is from 1 to 3600.
- Display all ARP entries learned via gratuitous ARP.

  EXEC Privilege mode

  ```
  show arp retries
  ```

ICMP

For diagnostics, the internet control message protocol (ICMP) provides routing information to end stations by choosing the best route (ICMP redirect messages) or determining if a router is reachable (ICMP Echo or Echo Reply).

ICMP error messages inform the router of problems in a particular packet. These messages are sent only on unicast traffic.

Configuration Tasks for ICMP

The following lists the configuration tasks for ICMP.

- Enabling ICMP Unreachable Messages

For a complete listing of all commands related to ICMP, refer to the *Dell Networking OS Command Line Reference Guide*.

Enabling ICMP Unreachable Messages

By default, ICMP unreachable messages are disabled.
When enabled, ICMP unreachable messages are created and sent out all interfaces.

To disable and re-enable ICMP unreachable messages, use the following commands.

- To disable ICMP unreachable messages.

  INTERFACE mode

  ```
  no ip unreachable
  ```
- Set FTOS to create and send ICMP unreachable messages on the interface.

  INTERFACE mode

  ```
  ip unreachable
  ```

To view if ICMP unreachable messages are sent on the interface, use the `show config` command in INTERFACE mode. If it is not listed in the `show config` command output, it is enabled. Only non-default information is displayed in the `show config` command output.

UDP Helper

User datagram protocol (UDP) helper allows you to direct the forwarding IP/UDP broadcast traffic by creating special broadcast addresses and rewriting the destination IP address of packets to match those addresses.

Configure UDP Helper

Configuring Dell Networking OS to direct UDP broadcast is a two-step process:

1. Enable UDP helper and specify the UDP ports for which traffic is forwarded. Refer to Enabling UDP Helper.
2. Configure a broadcast address on interfaces that will receive UDP broadcast traffic. Refer to Configuring a Broadcast Address.

Important Points to Remember

- The existing `ip directed broadcast` command is rendered meaningless if you enable UDP helper on the same interface.
- The broadcast traffic rate should not exceed 200 packets per second when you enable UDP helper.
- You may specify a maximum of 16 UDP ports.
- UDP helper is compatible with IP helper (`ip helper-address`):
  - UDP broadcast traffic with port number 67 or 68 are unicast to the dynamic host configuration protocol (DHCP) server per the `ip helper-address` configuration whether or not the UDP port list contains those ports.
  - If the UDP port list contains ports 67 or 68, UDP broadcast traffic is forwarded on those ports.

Enabling UDP Helper

To enable UDP helper, use the following command.

- Enable UPD helper.

  `ip udp-helper udp-ports`

Example of Enabling UDP Helper and Using the UDP Helper `show` Command

```
Dell(conf-if-gi-1/1)#ip udp-helper udp-port 1000
Dell(conf-if-gi-1/1)#show config
!
interface GigabitEthernet 1/1
  ip address 2.1.1.1/24
  ip udp-helper udp-port 1000
  no shutdown
```

To view the interfaces and ports on which you enabled UDP helper, use the `show ip udp-helper` command from EXEC Privilege mode.

```
Dell#show ip udp-helper
--------------------------------------------------
Port UDP port list
--------------------------------------------------
Gi 1/1 1000
```

Configuring a Broadcast Address

To configure a broadcast address, use the following command.

- Configure a broadcast address on an interface.

  `ip udp-broadcast-address`

Examples of Configuring and Viewing a Broadcast Address

```
Dell(conf-if-vl-100)#ip udp-broadcast-address 1.1.255.255
Dell(conf-if-vl-100)#show config
!
interface Vlan 100
ip address 1.1.0.1/24
ip udp-broadcast-address 1.1.255.255
untagged GigabitEthernet 1/2
no shutdown
```

To view the configured broadcast address for an interface, use `show interfaces` command.

```
R1_E600(conf)#do show interfaces vlan 100
Vlan 100 is up, line protocol is down
Address is 00:01:e8:0d:b9:7a, Current address is 00:01:e8:0d:b9:7a
Interface index is 1107787876
Internet address is 1.1.0.1/24
IP UDP-Broadcast address is 1.1.255.255
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed auto
ARP type: ARPA, ARP Timeout 04:00:00
Last clearing of "show interface" counters 00:07:44
Queueing strategy: fifo
Input Statistics:
      0 packets, 0 bytes
Time since last interface status change: 00:07:44
```

Configurations Using UDP Helper

When you enable UDP helper and the destination IP address of an incoming packet is a broadcast address, Dell Networking OS suppresses the destination address of the packet.

The following sections describe various configurations that employ UDP helper to direct broadcasts.

- UDP Helper with Broadcast-All Addresses
- UDP Helper with Subnet Broadcast Addresses
- UDP Helper with Configured Broadcast Addresses
- UDP Helper with No Configured Broadcast Addresses

UDP Helper with Broadcast-All Addresses

When the destination IP address of an incoming packet is the IP broadcast address, Dell Networking OS rewrites the address to match the configured broadcast address.

In the following illustration:

1. Packet 1 is dropped at ingress if you did not configure UDP helper address.
2. If you enable UDP helper (using the `ip udp-helper udp-port` command), and the UDP destination port of the packet matches the UDP port configured, the system changes the destination

address to the configured broadcast 1.1.255.255 and routes the packet to VLANs 100 and 101. If you do not configure an IP broadcast address (using the `ip udp-broadcast-address` command) on VLANs 100 or 101, the packet is forwarded using the original destination IP address 255.255.255.255.

Packet 2, sent from a host on VLAN 101 has a broadcast MAC address and IP address. In this case:

1. It is flooded on VLAN 101 without changing the destination address because the forwarding process is Layer 2.
2. If you enabled UDP helper, the system changes the destination IP address to the configured broadcast address 1.1.255.255 and forwards the packet to VLAN 100.
3. Packet 2 is also forwarded to the ingress interface with an unchanged destination address because it does not have broadcast address configured.

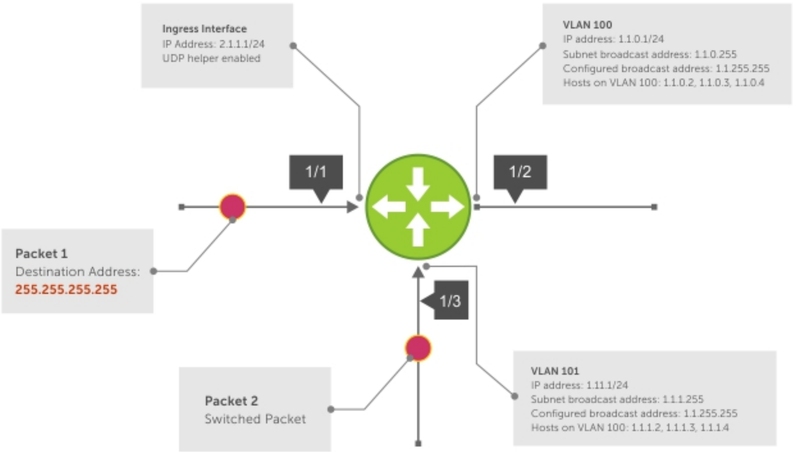

Ingress Interface
IP Address: 2.1.1.1/24
UDP helper enabled
VLAN 100
IP address: 1.1.0.1/24
Subnet broadcast address: 1.1.0.255
Configured broadcast address: 1.1.255.255
Hosts on VLAN 100: 1.1.0.2, 1.1.0.3, 1.1.0.4
1/1
1/2
Packet 1
Destination Address:
255.255.255.255
1/3
Packet 2
Switched Packet
VLAN 101
IP address: 1.11.1/24
Subnet broadcast address: 1.1.1.255
Configured broadcast address: 1.1.255.255
Hosts on VLAN 100: 1.1.1.2, 1.1.1.3, 1.1.1.4

Figure 44. UDP Helper with Broadcast-All Addresses

UDP Helper with Subnet Broadcast Addresses

When the destination IP address of an incoming packet matches the subnet broadcast address of any interface, the system changes the address to the configured broadcast address and sends it to matching interface.

In the following illustration, Packet 1 has the destination IP address 1.1.1.255, which matches the subnet broadcast address of VLAN 101. If you configured UDP helper and the packet matches the specified UDP port, the system changes the address to the configured IP broadcast address and floods the packet on VLAN 101.

Packet 2 is sent from the host on VLAN 101. It has a broadcast MAC address and a destination IP address of 1.1.1.255. In this case, it is flooded on VLAN 101 in its original condition as the forwarding process is Layer 2.

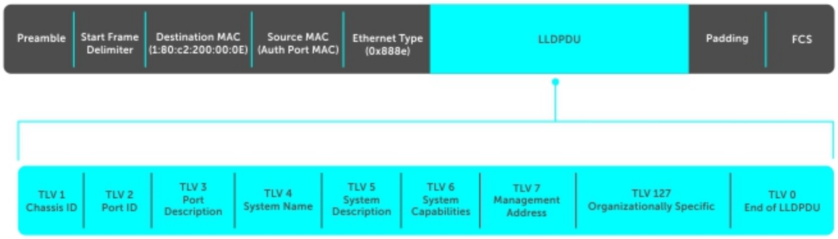

Preamble
Start Frame Delimiter
Destination MAC (1:80:c2:200:00:0E)
Source MAC (Auth Port MAC)
Ethernet Type (0x888e)
LLDPDU
Padding
FCS
TLV 1 Chassis ID
TLV 2 Port ID
TLV 3 Port Description
TLV 4 System Name
TLV 5 System Description
TLV 6 System Capabilities
TLV 7 Management Address
TLV 127 Organizationally Specific
TLV 0 End of LLDPDU

Figure 45. UDP Helper with Subnet Broadcast Addresses

UDP Helper with Configured Broadcast Addresses

Incoming packets with a destination IP address matching the configured broadcast address of any interface are forwarded to the matching interfaces.

In the following illustration, Packet 1 has a destination IP address that matches the configured broadcast address of VLAN 100 and 101. If you enabled UDP helper and the UDP port number matches, the packet is flooded on both VLANs with an unchanged destination address.

Packet 2 is sent from a host on VLAN 101. It has broadcast MAC address and a destination IP address that matches the configured broadcast address on VLAN 101. In this case, Packet 2 is flooded on VLAN 101 with the destination address unchanged because the forwarding process is Layer 2. If you enabled UDP helper, the packet is flooded on VLAN 100 as well.

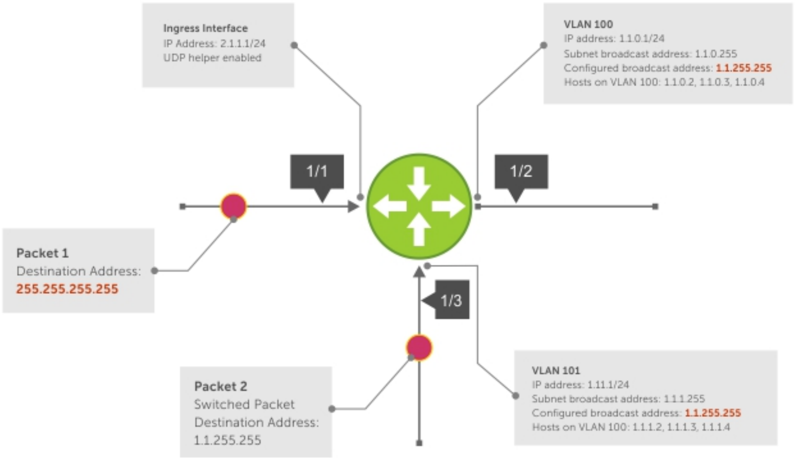

Ingress Interface
IP Address: 2.1.1.1/24
UDP helper enabled
VLAN 100
IP address: 1.1.0.1/24
Subnet broadcast address: 1.1.0.255
Configured broadcast address: 1.1.255.255
Hosts on VLAN 100: 1.1.0.2, 1.1.0.3, 1.1.0.4
1/1
1/2
Packet 1
Destination Address:
255.255.255.255
1/3
Packet 2
Switched Packet
Destination Address:
1.1.255.255
VLAN 101
IP address: 1.11.1/24
Subnet broadcast address: 1.1.1.255
Configured broadcast address: 1.1.255.255
Hosts on VLAN 100: 1.1.1.2, 1.1.1.3, 1.1.1.4

Figure 46. UDP Helper with Configured Broadcast Addresses

UDP Helper with No Configured Broadcast Addresses

The following describes UDP helper with no broadcast addresses configured.

- If the incoming packet has a broadcast destination IP address, the unaltered packet is routed to all Layer 3 interfaces.

- If the Incoming packet has a destination IP address that matches the subnet broadcast address of any interface, the unaltered packet is routed to the matching interfaces.

Troubleshooting UDP Helper

To display debugging information for troubleshooting, use the `debug ip udp-helper` command.

Example of the `debug ip udp-helper` Command

```
Dell(conf)# debug ip udp-helper
01:20:22: Pkt rcvd on Gi 5/0 with IP DA (0xffffffff) will be sent on Gi 5/1 Gi
5/2 Vlan 3
01:44:54: Pkt rcvd on Gi 7/0 is handed over for DHCP processing.
```

When using the IP helper and UDP helper on the same interface, use the `debug ip dhcp` command.

Example Output from the `debug ip dhcp` Command

```
Packet 0.0.0.0:68 -> 255.255.255.255:67 TTL 128

2005-11-05 11:59:35 %RELAY-I-PACKET, BOOTP REQUEST (Unicast) received at
interface
172.21.50.193 BOOTP Request, XID = 0x9265f901, secs = 0 hwaddr = 00:02:2D:8D:
46:DC,
giaddr = 0.0.0.0, hops = 2

2005-11-05 11:59:35 %RELAY-I-BOOTREQUEST, Forwarded BOOTREQUEST for 00:02:2D:8D:
46:DC
to 137.138.17.6

2005-11-05 11:59:36 %RELAY-I-PACKET, BOOTP REPLY (Unicast) received at interface
194.12.129.98 BOOTP Reply, XID = 0x9265f901, secs = 0 hwaddr = 00:02:2D:8D:
46:DC,
giaddr = 172.21.50.193, hops = 2

2005-07-05 11:59:36 %RELAY-I-BOOTREPLY, Forwarded BOOTREPLY for 00:02:2D:8D:
46:DC to
128.141.128.90 Packet 0.0.0.0:68 -> 255.255.255.255:67 TTL 128
```

21

IPv6 Routing

Internet protocol version 6 (IPv6) routing is supported on the S6000 platform.

NOTE: The IPv6 basic commands are supported on all platforms. However, not all features are supported on all platforms, nor for all releases. To determine the Dell Networking Operating System (OS) version supporting which features and platforms, refer to Implementing IPv6 with Dell Networking OS.

IPv6 is the successor to IPv4. Due to the rapid growth in internet users and IP addresses, IPv4 is reaching its maximum usage. IPv6 will eventually replace IPv4 usage to allow for the constant expansion.

This chapter provides a brief description of the differences between IPv4 and IPv6, and the Dell Networking support of IPv6. This chapter is not intended to be a comprehensive description of IPv6.

Protocol Overview

IPv6 is an evolution of IPv4. IPv6 is generally installed as an upgrade in devices and operating systems. Most new devices and operating systems support both IPv4 and IPv6.

Some key changes in IPv6 are:

- Extended address space
- Stateless autoconfiguration
- Header format simplification
- Improved support for options and extensions

Extended Address Space

The address format is extended from 32 bits to 128 bits. This not only provides room for all anticipated needs, it allows for the use of a hierarchical address space structure to optimize global addressing.

Stateless Autoconfiguration

When a booting device comes up in IPv6 and asks for its network prefix, the device can get the prefix (or prefixes) from an IPv6 router on its link. It can then autoconfigure one or more global IPv6 addresses by using either the MAC address or a private random number to build its unique IPv6 address.

Stateless autoconfiguration uses three mechanisms for IPv6 address configuration:

- **Prefix Advertisement** — Routers use “Router Advertisement” messages to announce the network prefix. Hosts then use their interface-identifier MAC address to generate their own valid IPv6 address.
- **Duplicate Address Detection (DAD)** — Before configuring its IPv6 address, an IPv6 host node device checks whether that address is used anywhere on the network using this mechanism.
- **Prefix Renumbering** — Useful in transparent renumbering of hosts in the network when an organization changes its service provider.

NOTE: As an alternative to stateless autoconfiguration, network hosts can obtain their IPv6 addresses using the dynamic host control protocol (DHCP) servers via stateful auto-configuration.

NOTE: Dell Networking OS provides the flexibility to add prefixes on Router Advertisements (RA) to advertise responses to Router Solicitations (RS). By default, RA response messages are sent when an RS message is received.

Dell Networking OS manipulation of IPv6 stateless autoconfiguration supports the router side only. Neighbor discovery (ND) messages are advertised so the neighbor can use this information to auto-configure its address. However, received ND messages are not used to create an IPv6 address.

NOTE: Inconsistencies in router advertisement values between routers are logged per RFC 4861. The values checked for consistency include:

- Cur Hop limit
- M and O flags
- Reachable time
- Retrans timer
- MTU options
- Preferred and valid lifetime values for the same prefix

Only management ports support stateless auto-configuration as a host.

The router redirect functionality in the neighbor discovery protocol (NDP) is similar to IPv4 router redirect messages. NDP uses ICMPv6 redirect messages (Type 137) to inform nodes that a better router exists on the link.

IPv6 Headers

The IPv6 header has a fixed length of 40 bytes. This fixed length provides 16 bytes each for source and destination information and 8 bytes for general header information.

The IPv6 header includes the following fields:

- Version (4 bits)
- Traffic Class (8 bits)
- Payload Length (16 bits)
- Next Header (8 bits)
- Hop Limit (8 bits)
- Source Address (128 bits)
- Destination Address (128 bits)

IPv6 provides for extension headers. Extension headers are used only if necessary. There can be no extension headers, one extension header or more than one extension header in an IPv6 packet. Extension headers are defined in the Next Header field of the preceding IPv6 header.

Longest Prefix Match (LPM) Table and IPv6 /65 – /128 support

Two partitions are available.

- Partition I with IPv6 /65 – /128 route prefix. Doesn’t support IPv4 entries in the current release.
- Partition II with IPv6 0/0 – /64 route prefix and IPv4 0/0 -0/32 route prefix entries.
- Number of entries in Partition II will be reduced based on the number of entries configured in Partition I.

- Partitioning will be applied well before the system initialization. This will be done using the NVRAM.
- Dell Networking OS provides CLI for enabling the partition. Configuration will be stored in NVRAM when the operator saves the configuration.
- Partition will take effect only after the switch reboot. During the reboot Dell Networking OS reads the partition configuration from NVRAM and uses the same for partitioning the LPM.
- A command has been introduced to partition the LPM to support provisioning of IPv6 /65 to /128 route prefixes.

To support /65 – /128 IPv6 route prefix entries, Dell Networking OS needs to be programmed with /65 - /128 bit IPv6 support on S6000 platform. The number of entries as well needs to be explicitly programmed. This number can be1K, 2K, or 3K granularity.

On the S6000 platform, for IPv6 /65 to /128 will consume the same storage banks which is used by the L3_DEFIP table. Once the IPv6 128 bit is enabled, number of entries in L3_DEFIP will be reduced.

LPM partitioning will take effect after reboot of the box. This is because the SDK does the LPM partitioning during the chip initialization.

The longest prefix match (LPM) table on the S6000 platform supports different types of prefixes for IPv6 and IPv4. The route table, also called the LPM table, is divided into the following three logical tables:

1. IPv4 32-bit LPM table (Holds IPv4 Prefixes)
2. IPv6 64-bit LPM table (Holds IPv6 Prefixes less than /65 Prefix Length)
3. IPv6 128-bit LPM table (Holds IPv6 Prefixes greater than /64 Prefix Length)

The LPM table, which is 8K in size, is a dedicated table. It comprises eight ternary content addressable memory (CAM) blocks, with each block being 1K in size. The table can contain 16KIPv4 route entries or 8K IPv6 route entries (less than /65 prefix-length) or 3K IPv6 route entries (greater than /64 prefix-length). You can configure the LPM table with one of the following partitions to support the IPv4 and IPv6 prefix route entries:

- Partition 1: IPv6 128-bit LPM entries can be stored in this partition. IPv4 and 64-bit IPv6 entries cannot be saved in this partition.
- Partition 2: IPv4 LPM and 64-bit IPv6 LPM entries can be stored in this partition.

The platforms (S6000) useS presently only IPv6 /0 – 0/64 prefix route entries. Support for /0 – /128 IPv6 prefix route entries is available, although they are not utilized. A total of eight pools or regions are present with each region containing 1024 210-bit entries (supports up to 0/64 prefix). To support up to /128 prefixes, you must use 2 banks (410-bit entries). It is necessary to partition the LPM.

The optimized booting functionality does not use Openflow and therefore SDN support is not available. LPM partitioning might have a slight impact on the number of SDN-programmed L3 entries because the LPM space becomes reduced.

IPv6 Header Fields

The 40 bytes of the IPv6 header are ordered, as shown in the following illustration.

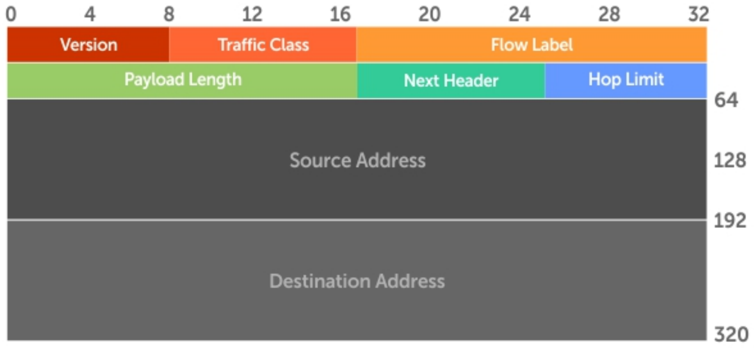


Figure 47. IPv6 Header Fields

Version (4 bits)

The Version field always contains the number 6, referring to the packet's IP version.

Traffic Class (8 bits)

The Traffic Class field deals with any data that needs special handling. These bits define the packet priority and are defined by the packet Source. Sending and forwarding routers use this field to identify different IPv6 classes and priorities. Routers understand the priority settings and handle them appropriately during conditions of congestion.

Flow Label (20 bits)

The Flow Label field identifies packets requiring special treatment in order to manage real-time data traffic.

The sending router can label sequences of IPv6 packets so that forwarding routers can process packets within the same flow without needing to reprocess each packet's header separately.

NOTE: All packets in the flow must have the same source and destination addresses.

Payload Length (16 bits)

The Payload Length field specifies the packet payload. This is the length of the data following the IPv6 header. IPv6 Payload Length only includes the data following the header, not the header itself.

The Payload Length limit of 2 bytes requires that the maximum packet payload be 64 KB. However, the Jumbogram option type Extension header supports larger packet sizes when required.

Next Header (8 bits)

The Next Header field identifies the next header's type. If an Extension header is used, this field contains the type of Extension header (as shown in the following table). If the next header is a transmission control protocol (TCP) or user datagram protocol (UDP) header, the value in this field is the same as for IPv4. The Extension header is located between the IP header and the TCP or UDP header.

The following lists the Next Header field values.

Value	Description
0	Hop-by-Hop option header
4	IPv4
6	TCP
8	Exterior Gateway Protocol (EGP)
41	IPv6
43	Routing header
44	Fragmentation header
50	Encrypted Security
51	Authentication header
59	No Next Header
60	Destinations option header

NOTE: This table is not a comprehensive list of Next Header field values. For a complete and current listing, refer to the Internet Assigned Numbers Authority (IANA) web page at .

Hop Limit (8 bits)

The Hop Limit field shows the number of hops remaining for packet processing. In IPv4, this is known as the Time to Live (TTL) field and uses seconds rather than hops.

Each time the packet moves through a forwarding router, this field decrements by 1. If a router receives a packet with a Hop Limit of 1, it decrements it to 0 (zero). The router discards the packet and sends an ICMPv6 message back to the sending router indicating that the Hop Limit was exceeded in transit.

Source Address (128 bits)

The Source Address field contains the IPv6 address for the packet originator.

Destination Address (128 bits)

The Destination Address field contains the intended recipient's IPv6 address. This can be either the ultimate destination or the address of the next hop router.

Extension Header Fields

Extension headers are used only when necessary. Due to the streamlined nature of the IPv6 header, adding extension headers do not severely impact performance. Each Extension headers's lengths vary, but they are always a multiple of 8 bytes.

Each extension header is identified by the Next Header field in the IPv6 header that precedes it. Extension headers are viewed only by the destination router identified in the Destination Address field. If the Destination Address is a multicast address, the Extension headers are examined by all the routers in that multicast group.

However, if the Destination Address is a Hop-by-Hop options header, the Extension header is examined by every forwarding router along the packet's route. The Hop-by-Hop options header must immediately follow the IPv6 header, and is noted by the value 0 (zero) in the Next Header field.

Extension headers are processed in the order in which they appear in the packet header.

Hop-by-Hop Options Header

The Hop-by-Hop options header contains information that is examined by every router along the packet's path. It follows the IPv6 header and is designated by the Next Header value 0 (zero).

When a Hop-by-Hop Options header is not included, the router knows that it does not have to process any router specific information and immediately processes the packet to its final destination.

When a Hop-by-Hop Options header is present, the router only needs this extension header and does not need to take the time to view further into the packet.

The Hop-by-Hop Options header contains:

- Next Header (1 byte)

This field identifies the type of header following the Hop-by-Hop Options header and uses the same values.

- Header Extension Length (1 byte)

This field identifies the length of the Hop-by-Hop Options header in 8-byte units, but does not include the first 8 bytes. Consequently, if the header is less than 8 bytes, the value is 0 (zero).

- Options (size varies)

This field can contain one or more options. The first byte if the field identifies the Option type, and directs the router how to handle the option.

00	Skip and continue processing.
01	Discard the packet.
10	Discard the packet and send an ICMP Parameter Problem Code 2 message to the packet's Source IP Address identifying the unknown option type.
11	Discard the packet and send an ICMP Parameter Problem, Code 2 message to the packet's Source IP Address only if the Destination IP Address is not a multicast address.

The second byte contains the Option Data Length.

The third byte specifies whether the information can change en route to the destination. The value is 1 if it can change; the value is 0 if it cannot change.

Addressing

IPv6 addresses are normally written as eight groups of four hexadecimal digits, where each group is separated by a colon (:).

For example, 2001:0db8:0000:0000:0000:0000:1428:57ab is a valid IPv6 address. If one or more four-digit group(s) is 0000, the zeros may be omitted and replaced with two colons(::). For example, 2001:0db8:0000:0000:0000:0000:1428:57ab can be shortened to 2001:0db8::1428:57ab. Only one set of double colons is supported in a single address. Any number of consecutive 0000 groups may be reduced to two colons, as long as there is only one double colon used in an address. Leading and/or trailing zeros in a group can also be omitted (as in ::1 for localhost, 1:: for network addresses and :: for unspecified addresses).

All the addresses in the following list are all valid and equivalent.

- 2001:0db8:0000:0000:0000:0000:1428:57ab
- 2001:0db8:0000:0000:0000::1428:57ab
- 2001:0db8:0:0:0:0:1428:57ab
- 2001:0db8:0:0::1428:57ab
- 2001:0db8::1428:57ab
- 2001:db8::1428:57ab

IPv6 networks are written using classless inter-domain routing (CIDR) notation. An IPv6 network (or subnet) is a contiguous group of IPv6 addresses the size of which must be a power of two; the initial bits of addresses, which are identical for all hosts in the network, are called the network's prefix.

A network is denoted by the first address in the network and the size in bits of the prefix (in decimal), separated with a slash. Because a single host is seen as a network with a 128-bit prefix, host addresses may be written with a following /128.

For example, 2001:0db8:1234::/48 stands for the network with addresses 2001:0db8:1234:0000:0000:0000:0000:0000 through 2001:0db8:1234:ffff:ffff:ffff:ffff:ffff.

Link-local Addresses

Link-local addresses, starting with fe80:, are assigned only in the local link area.

The addresses are generated usually automatically by the operating system's IP layer for each network interface. This provides instant automatic network connectivity for any IPv6 host and means that if several hosts connect to a common hub or switch, they have an instant communication path via their link-local IPv6 address.

Link-local addresses cannot be routed to the public Internet.

Static and Dynamic Addressing

Static IPv6 addresses are manually assigned to a computer by an administrator.

Dynamic IPv6 addresses are assigned either randomly or by a server using dynamic host configuration protocol (DHCP). Even though IPv6 addresses assigned using DHCP may stay the same for long periods of time, they can change. In some cases, a network administrator may implement dynamically assigned static IPv6 addresses. In this case, a DHCP server is used, but it is specifically configured to always assign the same IPv6 address to a particular computer, and never to assign that IP address to another computer. This allows static IPv6 addresses to be configured in one place, without having to specifically configure each computer on the network in a different way.

In IPv6, every interface, whether using static or dynamic address assignments, also receives a local-link address automatically in the fe80::/64 subnet.

Implementing IPv6 with Dell Networking OS

Dell Networking OS supports both IPv4 and IPv6 and both may be used simultaneously in your system.

The following table lists the Dell Networking OS version in which an IPv6 feature became available for each platform. The sections following the table give greater detail about the feature.

Feature and Functionality		Documentation and Chapter Location
	S6000	
Basic IPv6 Commands	S6000	IPv6 Basic Commands in the *Dell Networking OS Command Line Interface Reference Guide*.
IPv6 Basic Addressing		
IPv6 address types: Unicast	8.3.11	Extended Address Space
IPv6 neighbor discovery	8.3.11	IPv6 Neighbor Discovery
IPv6 stateless autoconfiguration	8.3.11	Stateless Autoconfiguration
IPv6 MTU path discovery	8.3.11	Path MTU Discovery
IPv6 ICMPv6	8.3.11	ICMPv6
IPv6 ping	8.3.11	ICMPv6
IPv6 traceroute	8.3.11	ICMPv6
IPv6 SNMP	8.3.11	
IPv6 Routing		
Static routing	8.3.11	Assigning a Static IPv6 Route
Route redistribution	8.3.11	OSPF, IS-IS, and IPv6 BGP chapters in the *Dell Networking OS Command Line Reference Guide*.
Multiprotocol BGP extensions for IPv6	8.3.11	IPv6 BGP in the *Dell Networking OS Command Line Reference Guide*.
IPv6 BGP MD5 Authentication	8.3.11	IPv6 BGP in the *Dell Networking OS Command Line Reference Guide*.

Feature and Functionality		Documentation and Chapter Location
	S6000	
IS-IS for IPv6	8.3.11	Intermediate System to Intermediate System IPv6 IS-IS in the *Dell Networking OS Command Line Reference Guide*.
IS-IS for IPv6 support for redistribution	8.3.11	Intermediate System to Intermediate System IPv6 IS-IS in the *Dell Networking OS Command Line Reference Guide*.
ISIS for IPv6 support for distribute lists and administrative distance	8.3.11	Intermediate System to Intermediate System IPv6 IS-IS in the *Dell Networking OS Command Line Reference Guide*.
OSPF for IPv6 (OSPFv3)	8.3.11	OSPFv3 in the *Dell Networking OS Command Line Reference Guide*.
Equal Cost Multipath for IPv6	8.3.11	
IPv6 Services and Management		
Telnet client over IPv6 (outbound Telnet)	8.3.11	Configuring Telnet with IPv6 Control and Monitoring in the *Dell Networking OS Command Line Reference Guide*.
Telnet server over IPv6 (inbound Telnet)	8.3.11	Configuring Telnet with IPv6 Control and Monitoring in the *Dell Networking OS Command Line Reference Guide*.
Secure Shell (SSH) client support over IPv6	8.3.11	Secure Shell (SSH) Over an IPv6 Transport

Feature and Functionality		Documentation and Chapter Location
	S6000	
(outbound SSH) Layer 3 only		
Secure Shell (SSH) server support over IPv6 (inbound SSH) Layer 3 only	8.3.11	Secure Shell (SSH) Over an IPv6 Transport
IPv6 Access Control Lists	8.3.11	IPv6 Access Control Lists in the *Dell Networking OS Command Line Reference Guide*.
IPv6 Multicast		
MLDv1/v2	N/A	IPv6 PIM in the *Dell Networking OS Command Line Reference Guide*.

ICMPv6

ICMPv6 is supported on the S6000 platform.

ICMP for IPv6 combines the roles of ICMP, IGMP and ARP in IPv4. Like IPv4, it provides functions for reporting delivery and forwarding errors, and provides a simple echo service for troubleshooting. The Dell Networking OS implementation of ICMPv6 is based on RFC 4443.

Generally, ICMPv6 uses two message types:

- Error reporting messages indicate when the forwarding or delivery of the packet failed at the destination or intermediate node. These messages include Destination Unreachable, Packet Too Big, Time Exceeded and Parameter Problem messages.
- Informational messages provide diagnostic functions and additional host functions, such as Neighbor Discovery and Multicast Listener Discovery. These messages also include Echo Request and Echo Reply messages.

The Dell Networking OS `ping` and `traceroute` commands extend to support IPv6 addresses. These commands use ICMPv6 Type-2 messages.

Path MTU Discovery

IPv6 path maximum transmission unit (MTU) discovery is supported on the S6000 platform.

Path MTU, in accordance with RFC 1981, defines the largest packet size that can traverse a transmission path without suffering fragmentation. Path MTU for IPv6 uses ICMPv6 Type-2 messages to discover the largest MTU along the path from source to destination and avoid the need to fragment the packet.

The recommended MTU for IPv6 is 1280. Greater MTU settings increase processing efficiency because each packet carries more data while protocol overheads (for example, headers) or underlying per-packet delays remain fixed.

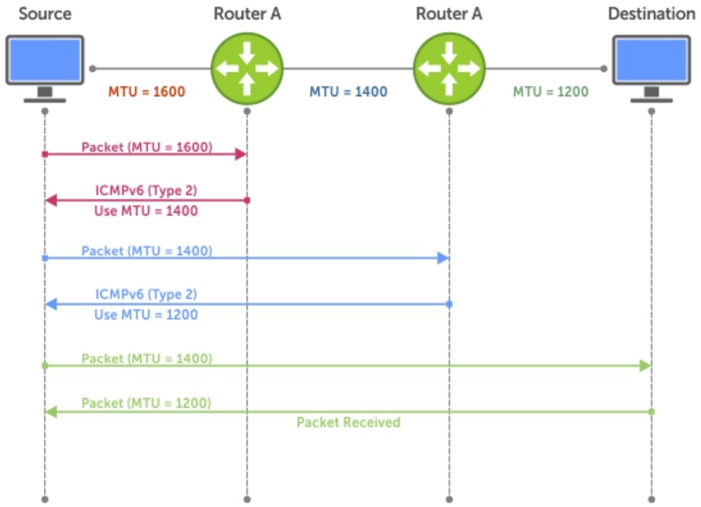

Source
Router A
Router A
Destination
MTU = 1600
MTU = 1400
MTU = 1200
Packet (MTU = 1600)
ICMPv6 (Type 2)
Use MTU = 1400
Packet (MTU = 1400)
ICMPv6 (Type 2)
Use MTU = 1200
Packet (MTU = 1400)
Packet (MTU = 1200)
Packet Received

Figure 48. Path MTU Discovery Process

IPv6 Neighbor Discovery

IPv6 neighbor discovery protocol (NDP) is supported on the S6000 platform.

NDP is a top-level protocol for neighbor discovery on an IPv6 network. In lieu of address resolution protocol (ARP), NDP uses “Neighbor Solicitation” and “Neighbor Advertisement” ICMPv6 messages for determining relationships between neighboring nodes. Using these messages, an IPv6 device learns the link-layer addresses for neighbors known to reside on attached links, quickly purging cached values that become invalid.

NOTE: If a neighboring node does not have an IPv6 address assigned, it must be manually pinged to allow the IPv6 device to determine the relationship of the neighboring node.

NOTE: To avoid problems with network discovery, Dell Networking recommends configuring the static route last or assigning an IPv6 address to the interface and assigning an address to the peer (the forwarding router’s address) less than 10 seconds apart.

With ARP, each node broadcasts ARP requests on the entire link. This approach causes unnecessary processing by uninterested nodes. With NDP, each node sends a request only to the intended destination via a multicast address with the unicast address used as the last 24 bits. Other hosts on the link do not participate in the process, greatly increasing network bandwidth efficiency.

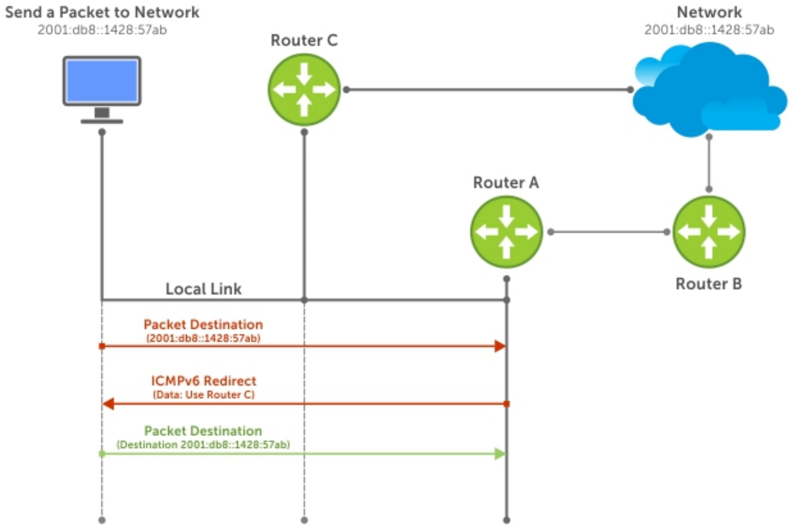

Send a Packet to Network
2001:db8::1428:57ab
Router C
Network
2001:db8::1428:57ab
Router A
Router B
Local Link
Packet Destination
(2001:db8::1428:57ab)
ICMPv6 Redirect
(Data: Use Router C)
Packet Destination
(Destination 2001:db8::1428:57ab)

Figure 49. NDP Router Redirect

IPv6 Neighbor Discovery of MTU Packets

You can set the MTU advertised through the RA packets to incoming routers, without altering the actual MTU setting on the interface.

The `ipv6 nd mtu` command sets the value advertised to routers. It does not set the actual MTU rate. For example, if you set `ipv6 nd mtu` to 1280, the interface still passes 1500-byte packets, if that is what is set with the `mtu` command.

Configuring the IPv6 Recursive DNS Server

You can configure up to four Recursive DNS Server (RDNSS) addresses to be distributed via IPv6 router advertisements to an IPv6 device, using the `ipv6 nd dns-server` *`ipv6-RDNSS-address`* `{`*`lifetime`* `| infinite}` command in INTERFACE CONFIG mode.

The lifetime parameter configures the amount of time the IPv6 host can use the IPv6 RDNSS address for name resolution. The lifetime range is `0` to `4294967295` seconds. When the maximum lifetime value, `4294967295`, or the `infinite` keyword is specified, the lifetime to use the RDNSS address does not expire. A value of `0` indicates to the host that the RDNSS address should not be used. You must specify a lifetime using the lifetime or infinite parameter.

The DNS server address does not allow the following:

- link local addresses
- loopback addresses
- prefix addresses
- multicast addresses
- invalid host addresses

If you specify this information in the IPv6 RDNSS configuration, a DNS error is displayed.

Example for Configuring an IPv6 Recursive DNS Server

The following example configures a RDNNS server with an IPv6 address of `1000::1` and a lifetime of `1` second.

```
Dell(conf-if-te-0/1)#ipv6 nd dns-server ?
X:X:X:X::X          Recursive DNS Server's (RDNSS) IPv6 address
Dell(conf-if-te-0/1)#ipv6 nd dns-server 1000::1 ?
<0-4294967295>       Max lifetime (sec) which RDNSS address may be used for
name resolution
infinite             Infinite lifetime (sec) which RDNSS address may be used
for name resolution

Dell(conf-if-te-0/1)#ipv6 nd dns-server 1000::1 1
```

Debugging IPv6 RDNSS Information Sent to the Host

To verify that the IPv6 RDNSS information sent to the host is configured correctly, use the `debug ipv6 nd` command in EXEC Privilege mode.

Example of Debugging IPv6 RDNSS Information Sent to the Host

The following example debugs IPv6 RDNSS information sent to the host.

```
Dell(conf-if-te-0/1)#do debug ipv6 nd tengigabitethernet 0/1
ICMPv6 Neighbor Discovery packet debugging is on for tengigabitethernet 0/1
Dell(conf-if-te-0/1)#00:13:02 : : cp-ICMPV6-ND: Sending RA on Te 0/1
            current hop limit=64, flags: M-, O-,
           router lifetime=1800 sec, reachable time=0 ms, retransmit time=0 ms
            SLLA=00:01:e8:8b:75:70
            prefix=1212::/64 on-link autoconfig
            valid lifetime=2592000 sec, preferred lifetime=604800 sec
            dns-server=1000::0001, lifetime=1 sec
            dns-server=3000::0001, lifetime=1 sec
            dns-server=2000::0001, lifetime=0 sec
```

The last 3 lines indicate that the IPv6 RDNSS information was configured correctly.

```
dns-server=1000::0001, lifetime=1 sec
dns-server=3000::0001, lifetime=1 sec
dns-server=2000::0001, lifetime=0 sec
```

If the DNS server information is not displayed, verify that the IPv6 recursive DNS server configuration was configured on the correct interface.

Displaying IPv6 RDNSS Information

To display IPv6 interface information, including IPv6 RDNSS information, use the `show ipv6 interface` command in EXEC or EXEC Privilege mode.

Examples of Displaying IPv6 RDNSS Information

The following example displays IPv6 RDNSS information. The output in the last 3 lines indicates that the IPv6 RDNSS was correctly configured on interface `te 0/1`.

```
Dell#show ipv6 interface te 0/1
TenGigabitEthernet 0/1 is up, line protocol is up
  IPV6 is enabled
  Link Local address: fe80::201:e8ff:fe8b:7570
  Global Unicast address(es):
    1212::12, subnet is 1212::/64 (MANUAL)
     Remaining lifetime: infinite
  Global Anycast address(es):
```

```
  Joined Group address(es):
    ff02::1
    ff02::2
    ff02::1:ff00:12
    ff02::1:ff8b:7570
  ND MTU is 0
  ICMP redirects are not sent
  DAD is enabled, number of DAD attempts: 3
  ND reachable time is 20120 milliseconds
  ND base reachable time is 30000 milliseconds
  ND advertised reachable time is 0 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 198 to 600 seconds
  ND router advertisements live for 1800 seconds
  ND advertised hop limit is 64
  IPv6 hop limit for originated packets is 64
  ND dns-server address is 1000::1 with lifetime of 1 seconds
  ND dns-server address is 3000::1 with lifetime of 1 seconds
  ND dns-server address is 2000::1 with lifetime of 0 seconds
```

To display IPv6 RDNSS information, use the `show configuration` command in INTERFACE CONFIG mode.

```
Dell(conf-if-te-0/1)#show configuration
```

The following example uses the `show configuration` command to display IPv6 RDNSS information.

```
!
interface TenGigabitEthernet 0/1
no ip address
ipv6 address 1212::12/64
ipv6 nd dns-server 1000::1 1
ipv6 nd dns-server 3000::1 1
ipv6 nd dns-server 2000::1 0
no shutdown
```

Secure Shell (SSH) Over an IPv6 Transport

IPv6 secure shell (SSH) is supported on the S6000 platform.

Dell Networking OS supports both inbound and outbound SSH sessions using IPv6 addressing. Inbound SSH supports accessing the system through the management interface as well as through a physical Layer 3 interface.

For SSH configuration details, refer to the *Security* chapter in the *Dell Networking OS Command Line Interface Reference Guide*.

Configuration Tasks for IPv6

The following are configuration tasks for the IPv6 protocol.

- Adjusting Your CAM-Profile
- Assigning an IPv6 Address to an Interface
- Assigning a Static IPv6 Route
- Configuring Telnet with IPv6
- SNMP over IPv6
- Showing IPv6 Information

- Clearing IPv6 Routes

Adjusting Your CAM-Profile

The `cam-acl` command is supported on the S6000 platform.
Although adjusting your CAM-profile is not a mandatory step, if you plan to implement IPv6 ACLs, adjust your CAM settings.

The CAM space is allotted in FP blocks. The total space allocated must equal 13 FP blocks. There are 16 FP blocks, but the System Flow requires three blocks that cannot be reallocated.

You must enter the `ipv6acl` allocation as a factor of 2 (2, 4, 6, 8, 10). All other profile allocations can use either even or odd-numbered ranges.

The default option sets the CAM Profile as follows:

- L3 ACL (ipv4acl): 6
- L2 ACL(l2acl): 5
- IPv6 L3 ACL (ipv6acl): 0
- L3 QoS (ipv4qos): 1
- L2 QoS (l2qos): 1

To have the changes take effect, save the new CAM settings to the startup-config (`write-mem` or `copy run start`) then reload the system for the new settings.

- Allocate space for IPV6 ACLs. Enter the CAM profile name then the allocated amount.

  CONFIGURATION mode

  ```
  cam-acl { ipv6acl }
  ```

  When not selecting the default option, enter all of the profiles listed and a range for each.

  The total space allocated must equal 13.

  The `ipv6acl` range must be a factor of 2.
- Show the current CAM settings.

  EXEC mode or EXEC Privilege mode

  ```
  show cam-acl
  ```
- Provides information on FP groups allocated for the egress acl.

  CONFIGURATION mode

  ```
  show cam-acl-egress
  ```

  Allocate at least one group for L2ACL and IPv4 ACL.

  The total number of groups is 4.

Assigning an IPv6 Address to an Interface

IPv6 addresses are supported on the S6000 platform.
Essentially, IPv6 is enabled in Dell Networking OS simply by assigning IPv6 addresses to individual router interfaces. You can use IPv6 and IPv4 together on a system, but be sure to differentiate that usage carefully. To assign an IPv6 address to an interface, use the `ipv6 address` command.

You can configure up to two IPv6 addresses on management interfaces, allowing required default router support on the management port that is acting as host, per RFC 4861. Data ports support more than two IPv6 addresses.

When you configure IPv6 addresses on multiple interfaces (the `ipv6 address` command) and verify the configuration (the `show ipv6 interfaces` command), the same link local (fe80) address is displayed for each IPv6 interface.

- Enter the IPv6 Address for the device.

  CONFIG-INTERFACE mode

  ```
  ipv6 address ipv6 address/mask
  ```

  - *`ipv6 address`*: x:x:x:x::x
  - *`mask`*: The prefix length is from 0 to 128

  NOTE: IPv6 addresses are normally written as eight groups of four hexadecimal digits. Separate each group by a colon (:). Omitting zeros is accepted as described in Addressing.

Assigning a Static IPv6 Route

IPv6 static routes are supported on the S6000 platform.
To configure IPv6 static routes, use the `ipv6 route` command.

NOTE: After you configure a static IPv6 route (the `ipv6 route` command) and configure the forwarding router's address (specified in the `ipv6 route` command) on a neighbor's interface, the IPv6 neighbor does not display in the `show ipv6 route` command output.

- Set up IPv6 static routes.

  CONFIGURATION mode

  ```
  ipv6 route prefix type {slot/port} forwarding router tag
  ```

  - *`prefix`*: IPv6 route prefix
  - *`type {slot/port}`*: interface type and slot/port
  - *`forwarding router`*: forwarding router's address
  - *`tag`*: route tag

  Enter the keyword `interface` then the type of interface and slot/port information:

  - For a 10/100/1000 Ethernet interface, enter the keyword `GigabitEthernet` then the slot/ port information.
  - For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/ port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
  - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.
  - For a loopback interface, enter the keyword `loopback` then the loopback number.
  - For a port-channel interface, enter the keywords `port-channel` then the port-channel number.
  - For a VLAN interface, enter the keyword `vlan` then the VLAN ID.
  - For a Null interface, enter the keyword `null` then the Null interface number.

Configuring Telnet with IPv6

IPv6 telnet is supported on the S6000 platform.
The Telnet client and server in Dell Networking OS supports IPv6 connections. You can establish a Telnet session directly to the router using an IPv6 Telnet client, or you can initiate an IPv6 Telnet connection from the router.

- Enter the IPv6 Address for the device.

  EXEC mode or EXEC Privileged mode

  `telnet *ipv6 address*`

  - *`ipv6 address`*: x:x:x:x::x
  - *`mask`*: prefix length is from 0 to 128.

NOTE: IPv6 addresses are normally written as eight groups of four hexadecimal digits, where each group is separated by a colon (:). Omitting zeros is accepted as described in Addressing.

SNMP over IPv6

The simple network management protocol (SNMP) is supported on the S6000 platform.

You can configure SNMP over IPv6 transport so that an IPv6 host can perform SNMP queries and receive SNMP notifications from a device running Dell Networking OS IPv6. The Dell Networking OS SNMP-server commands for IPv6 have been extended to support IPv6. For more information regarding SNMP commands, refer to the *SNMP* and *SYSLOG* chapters in the *Dell Networking OS Command Line Interface Reference Guide*.

- `snmp-server host`
- `snmp-server user ipv6`
- `snmp-server community ipv6`
- `snmp-server community access-list-name ipv6`
- `snmp-server group ipv6`
- `snmp-server group access-list-name ipv6`

Showing IPv6 Information

All of the following `show` commands are supported on the S6000 platform.
View specific IPv6 configuration with the following commands.

- List the IPv6 show options.

  EXEC mode or EXEC Privileged mode

  `show ipv6 ?`

Example of `show ipv6` Command Options

```
Dell#show ipv6 ?
accounting   IPv6 accounting information
cam          IPv6 CAM Entries
fib          IPv6 FIB Entries
interface    IPv6 interface information
mbgproutes   MBGP routing table
mld          MLD information
```

```
mroute       IPv6 multicast-routing table
neighbors    IPv6 neighbor information
ospf         OSPF information
pim          PIM V6 information
prefix-list  List IPv6 prefix lists
route        IPv6 routing information
rpf          RPF table
Dell#
```

Showing an IPv6 Interface

To view the IPv6 configuration for a specific interface, use the following command.

- Show the currently running configuration for the specified interface.

  EXEC mode

  `show ipv6 interface type {slot/port}`

  Enter the keyword `interface` then the type of interface and slot/port information:

  - For all brief summary of IPv6 status and configuration, enter the keyword `brief`.
  - For all IPv6 configured interfaces, enter the keyword `configured`.
  - For a 10/100/1000 Ethernet interface, enter the keyword `GigabitEthernet` then the slot/ port information.
  - For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/ port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
  - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.
  - For a loopback interface, enter the keyword `loopback` then the loopback number.
  - For a port-channel interface, enter the keywords `port-channel` then the port-channel number.
  - For a VLAN interface, enter the keyword `vlan` then the VLAN ID.

Example of the `show ipv6 interface` Command ()

```
Dell#show ipv6 int man 1/0
ManagementEthernet 1/0 is up, line protocol is up
  IPV6 is enabled
  Stateless address autoconfiguration is enabled
  Link Local address: fe80::201:e8ff:fe8b:386e
  Global Unicast address(es):
    Actual address is 400::201:e8ff:fe8b:386e, subnet is 400::/64
    Actual address is 412::201:e8ff:fe8b:386e, subnet is 412::/64
     Virtual-IP IPv6 address is not set
  Received Prefix(es):
    400::/64 onlink autoconfig
     Valid lifetime: 2592000, Preferred lifetime: 604800
     Advertised by: fe80::201:e8ff:fe8b:3166
  412::/64 onlink autoconfig
     Valid lifetime: 2592000, Preferred lifetime: 604800
     Advertised by: fe80::201:e8ff:fe8b:3166
  Global Anycast address(es):
  Joined Group address(es):
     ff02::1
     ff02::1:ff8b:386e
ND MTU is 0
ICMP redirects are not sent
DAD is enabled, number of DAD attempts: 3
ND reachable time is 32000 milliseconds
ND base reachable time is 30000 milliseconds
ND retransmit interval is 1000 milliseconds
ND hop limit is 64
```

Showing IPv6 Routes

To view the global IPv6 routing information, use the following command.

- Show IPv6 routing information for the specified route type.
  EXEC mode

  ```
  show ipv6 route type
  ```

  The following keywords are available:
  - To display information about a network, enter `ipv6 address` (X:X:X:X::X).
  - To display information about a host, enter `hostname`.
  - To display information about all IPv6 routes (including non-active routes), enter `all`.
  - To display information about all connected IPv6 routes, enter `connected`.
  - To display information about brief summary of all IPv6 routes, enter `summary`.
  - To display information about Border Gateway Protocol (BGP) routes, enter `bgp`.
  - To display information about ISO IS-IS routes, enter `isis`.
  - To display information about Open Shortest Path First (OSPF) routes, enter `ospf`.
  - To display information about Routing Information Protocol (RIP), enter `rip`.
  - To display information about static IPv6 routes, enter `static`.
  - To display information about an IPv6 Prefix lists, enter `list` and the prefix-list name.

Examples of the `show ipv6 route` Commands

The following example shows the `show ipv6 route summary` command.

```
Dell#show ipv6 route summary

Route Source Active Routes Non-active Routes
connected 5 0
static 0 0
Total 5 0
```

The following example shows the `show ipv6 route` command.

```
Dell#show ipv6 route
Codes: C - connected, L - local, S - static, R - RIP,
       B - BGP, IN - internal BGP, EX - external BGP,LO - Locally Originated,
       O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
       N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
       E2 - OSPF external type 2, i - IS-IS, L1 - IS-IS level-1,
       L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
Gateway of last resort is not set

   Destination  Dist/Metric,        Gateway, Last Change
-----------------------------------------------------
C  600::/64 [0/0]
      Direct, Te 0/24, 00:34:42
C  601::/64 [0/0]
      Direct, Te 0/24, 00:34:18
```

```
C  912::/64 [0/0]
      Direct, Lo 2, 00:02:33
O  IA 999::1/128 [110/2]
      via fe80::201:e8ff:fe8b:3166, Te 0/24, 00:01:30
L  fe80::/10 [0/0]
      Direct, Nu 0, 00:34:42
Dell#
```

The following example shows the `show ipv6 route static` command.

```
Dell#show ipv6 route static
Destination Dist/Metric, Gateway, Last Change
-----------------------------------------------------
  S         8888:9999:5555:6666:1111:2222::/96 [1/0]
               via         2222:2222:3333:3333::1, Gi 9/1, 00:03:16
  S         9999:9999:9999:9999::/64 [1/0]
               via 8888:9999:5555:6666:1111:2222:3333:4444, 00:03:16
```

Showing the Running-Configuration for an Interface

To view the configuration for any interface, use the following command.

- Show the currently running configuration for the specified interface.

  EXEC mode

  `show running-config interface` *`type`* `{`*`slot/port`*`}`

  Enter the keyword `interface` then the type of interface and slot/port information:

  - For a 10/100/1000 Ethernet interface, enter the keyword `GigabitEthernet` then the slot/ port information.
  - For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/ port information.
  - For the Management interface on the RPM, enter the keyword `ManagementEthernet` then the slot/port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
  - For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.

Example of the `show running-config interface` Command

```
Dell#show run int gi 2/2
!
interface GigabitEthernet 2/2
  no ip address
  ipv6 address 3:4:5:6::8/24
  shutdown
Dell#
```

Clearing IPv6 Routes

To clear routes from the IPv6 routing table, use the following command.

- Clear (refresh) all or a specific route from the IPv6 routing table.

  EXEC mode

  `clear ipv6 route {* |` *`ipv6 address prefix-length`*`}`

  - `*`: all routes.

  - *`ipv6 address`*: the format is x:x:x:x::x.
  - *`mask`*: the prefix length is from 0 to 128.

NOTE: IPv6 addresses are normally written as eight groups of four hexadecimal digits, where each group is separated by a colon (:). Omitting zeros is accepted as described in Addressing.

22

iSCSI Optimization

iSCSI optimization is supported on the S6000 platform.

This chapter describes how to configure internet small computer system interface (iSCSI) optimization, which enables quality-of-service (QoS) treatment for iSCSI traffic. The topics covered in this chapter include:

- iSCSI Optimization Overview
- Default iSCSI Optimization Values
- iSCSI Optimization Prerequisites
- Configuring iSCSI Optimization
- Displaying iSCSI Optimization Information

iSCSI Optimization Overview

iSCSI is a TCP/IP-based protocol for establishing and managing connections between IP-based storage devices and initiators in a storage area network (SAN).

iSCSI optimization enables the network switch to auto-detect Dell's iSCSI storage arrays and triggers a self-configuration of several key network configurations that enables optimization of the network for better storage traffic throughput. iSCSI is disabled by default on the platform and is not supported on the Z9000 platform.

iSCSI optimization also provides a means of monitoring iSCSI sessions and applying quality of service (QoS) policies on iSCSI traffic. When enabled, iSCSI optimization allows a switch to monitor (snoop) the establishment and termination of iSCSI connections. The switch uses the snooped information to detect iSCSI sessions and connections established through the switch.

iSCSI optimization allows you to reduce deployment time and management complexity in data centers. In a data center network, Dell EqualLogic and Compellent iSCSI storage arrays are connected to a converged Ethernet network using the data center bridging exchange protocol (DCBx) through stacked and/or non-stacked Ethernet switches.

iSCSI session monitoring over virtual link trunking (VLT) synchronizes the iSCSI session information between the VLT peers, allowing session information to be available in both the VLT peers. You can enable or disable iSCSI when you configure VLT.

iSCSI optimization functions as follows:

- Auto-detection of EqualLogic storage arrays — the switch detects any active EqualLogic array directly attached to its ports.
- Manual configuration to detect Compellent storage arrays where auto-detection is not supported.
- Automatic configuration of switch ports after detection of storage arrays.

- If you configure flow-control, iSCSI uses the current configuration. If you do not configure flow-control, iSCSI auto-configures flow control settings so that receive-only is enabled and transmit-only is disabled. .
- iSCSI monitoring sessions — the switch monitors and tracks active iSCSI sessions in connections on the switch, including port information and iSCSI session information.
- iSCSI QoS — A user-configured iSCSI class of service (CoS) profile is applied to all iSCSI traffic. Classifier rules are used to direct the iSCSI data traffic to queues that can be given preferential QoS treatment over other data passing through the switch. Preferential treatment helps to avoid session interruptions during times of congestion that would otherwise cause dropped iSCSI packets.
- iSCSI DCBx TLVs are supported.

The following illustration shows iSCSI optimization between servers and a storage array in which a stack of three switches connect installed servers (iSCSI initiators) to a storage array (iSCSI targets) in a SAN network. iSCSI optimization running on the master switch is configured to use dot1p priority-queue assignments to ensure that iSCSI traffic in these sessions receives priority treatment when forwarded on stacked switch hardware.

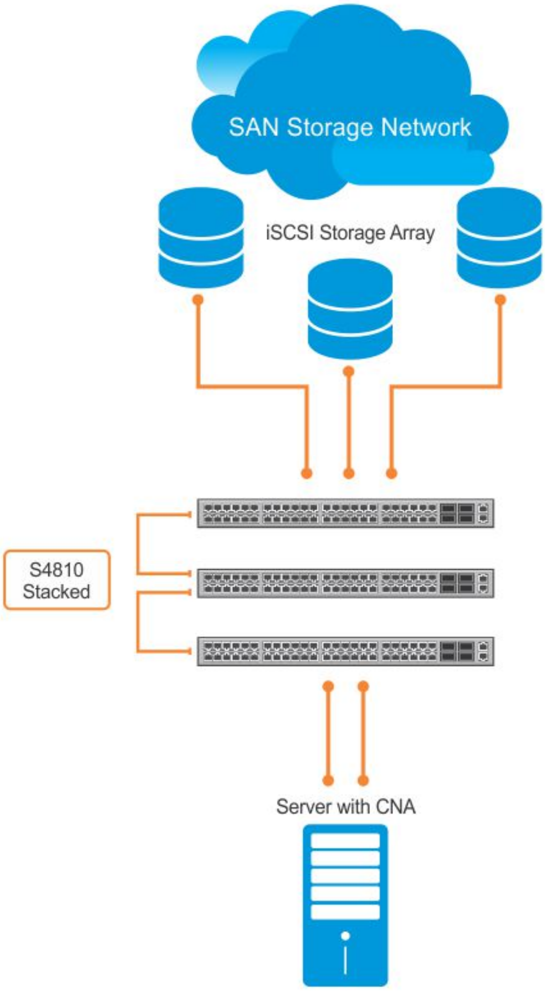


Figure 50. iSCSI Optimization Example

Monitoring iSCSI Traffic Flows

The switch snoops iSCSI session-establishment and termination packets by installing classifier rules that trap iSCSI protocol packets to the CPU for examination.

Devices that initiate iSCSI sessions usually use well-known TCP ports 3260 or 860 to contact targets. When you enable iSCSI optimization, by default the switch identifies IP packets to or from these ports as iSCSI traffic.

You can configure the switch to monitor traffic for additional port numbers or a combination of port number and target IP address, and you can remove the well-known port numbers from monitoring.

Application of Quality of Service to iSCSI Traffic Flows

You can configure iSCSI CoS mode. This mode controls whether CoS (dot1p priority) queue assignment and/or packet marking is performed on iSCSI traffic.

When you enable iSCSI CoS mode, the CoS policy is applied to iSCSI traffic. When you disable iSCSI CoS mode, iSCSI sessions and connections are still detected and displayed in the status tables, but no CoS policy is applied to iSCSI traffic.

You can configure whether the iSCSI optimization feature uses the VLAN priority or IP DSCP mapping to determine the traffic class queue. By default, iSCSI flows are assigned to dot1p priority 4. To map incoming iSCSI traffic on an interface to a dot1p priority-queue other than 4, use the `CoS dot1p-priority` command (refer to QoS dot1p Traffic Classification and Queue Assignment). Dell Networking recommends setting the CoS dot1p priority-queue to 0 (zero).

You can configure whether iSCSI frames are re-marked to contain the configured VLAN priority tag or IP DSCP when forwarded through the switch.

NOTE: On a switch in which a large proportion of traffic is iSCSI, CoS queue assignments may interfere with other network control-plane traffic, such as ARP or LACP. Balance preferential treatment of iSCSI traffic against the needs of other critical data in the network.

Information Monitored in iSCSI Traffic Flows

iSCSI optimization examines the following data in packets and uses the data to track the session and create the classifier entries that enable QoS treatment.

- Initiator's IP Address
- Target's IP Address
- ISID (Initiator defined session identifier)
- Initiator's IQN (iSCSI qualified name)
- Target's IQN
- Initiator's TCP Port
- Target's TCP Port
- Connection ID
- Aging
- Up Time

If no iSCSI traffic is detected for a session during a user-configurable aging period, the session data is cleared.

If more than 256 simultaneous sessions are logged continuously, the following message displays indicating the queue rate limit has been reached:

```
%STKUNIT2-M:CP %iSCSI-5-ISCSI_OPT_MAX_SESS_EXCEEDED: New iSCSI Session Ignored:
ISID -
400001370000 InitiatorName - iqn.1991-05.com.microsoft:dt-brcd-cna-2 TargetName
-
iqn.2001-05.com.equallogic:4-52aed6-b90d9446c-162466364804fa49-wj-v1 TSIH - 0"
```

NOTE: If you are using EqualLogic or Compellent storage arrays, more than 256 simultaneous iSCSI sessions are possible. However, iSCSI session monitoring is not capable of monitoring more than 256 simultaneous iSCSI sessions. If this number is exceeded, sessions may display as `unknown` in session monitoring output. Dell Networking recommends that you disable iSCSI session monitoring for EqualLogic and Compellent storage arrays or for installations with more than 256 simultaneous iSCSI sessions.

Only sessions the switch observes are learned; sessions flowing through an adjacent switch are not learned. Session monitoring learns sessions that actually flow through the switch, it does not learn all sessions in the entire topology.

After a switch is reloaded, any information exchanged during the initial handshake is not available. If the switch picks up the communication after reloading, it would detect a session was in progress but could not obtain complete information for it. Any incomplete information of this type would not be available in the `show` commands.

Detection and Auto-Configuration for Dell EqualLogic Arrays

The iSCSI optimization feature includes auto-provisioning support with the ability to detect directly connected Dell EqualLogic storage arrays and automatically reconfigure the switch to enhance storage traffic flows.

The switch uses the link layer discovery protocol (LLDP) to discover Dell EqualLogic devices on the network. LLDP is enabled by default. For more information about LLDP, refer to Link Layer Discovery Protocol (LLDP).

The following message displays the first time a Dell EqualLogic array is detected and describes the configuration changes that are automatically performed:

```
%STKUNIT0-M:CP %IFMGR-5-IFM_ISCSI_AUTO_CONFIG: This switch is being configured
for optimal
conditions to support iSCSI traffic which will cause some automatic
configuration to occur
including jumbo frames and flow-control on all ports; no storm control and
spanning-tree port
fast to be enabled on the port of detection.
```

The following syslog message is generated the first time an EqualLogic array is detected:

```
%STKUNIT0-M:CP %LLDP-5-LLDP_EQL_DETECTED: EqualLogic Storage Array detected on
interface Te 1/
43
```

- At the first detection of an EqualLogic array, an MTU of 12000 is enabled on all ports and port-channels (if it has not already been enabled).
- Spanning-tree portfast is enabled on the interface LLDP identifies.
- Unicast storm control is disabled on the interface LLDP identifies.

Configuring Detection and Ports for Dell Compellent Arrays

To configure a port connected to a Dell Compellent storage array, use the following command.

- Configure a port connected to a Dell Compellent storage array.

  INTERFACE Configuration mode

  ```
  iscsi profile-compellent
  ```

  The command configures a port for the best iSCSI traffic conditions.

The following message displays the first time you use the `iscsi profile-compellent` command to configure a port connected to a Dell Compellent storage array and describes the configuration changes that are automatically performed:

```
%STKUNIT0-M:CP %IFMGR-5-IFM_ISCSI_AUTO_CONFIG: This switch is being configured
for optimal
conditions to support iSCSI traffic which will cause some automatic
configuration to occur
including jumbo frames and flow-control on all ports; no storm control and
spanning-tree port
fast to be enabled on the port of detection.
```

After you execute the `iscsi profile-compellent` command, the following actions occur:

- Jumbo frame size is set to 1200 for all interfaces on all ports and port-channels, if it is not already enabled.
- Spanning-tree portfast is enabled on the interface.
- Unicast storm control is disabled on the interface.

Enter the `iscsi profile-compellent` command in INTERFACE Configuration mode; for example:

```
Dell(conf-if-te-o/50# iscsi profile-compellent
```

Synchronizing iSCSI Sessions Learned on VLT-Lags with VLT-Peer

The following behavior occurs during synchronization of iSCSI sessions.

- If the iSCSI login request packet is received on a port belonging to a VLT lag, the information is synced to the VLT peer and the connection is associated with this interface.
- Additional updates to connections (including aging updates) that are learnt on VLT lag members are synced to the peer.
- When receiving an iSCSI login request on a non-VLT interface followed by a response from a VLT interface, the session is not synced since it is initially learnt on a non-VLT interface through the request packet.
- The peer generates a new connection log that sees the login response packet. If the login response packet uses the ICL path, it is seen by both the peers, which in turn generate logs for this connection.

Enable and Disable iSCSI Optimization

The following describes enabling and disabling iSCSI optimizaiton.

NOTE: iSCSI monitoring is disabled by default. iSCSI auto-configuration and auto-detection is enabled by default.

If you enable iSCSI, flow control is automatically enabled on all interfaces. To disable flow control on all interfaces, use the `no flow control rx on tx off` command and save the configuration. To disable iSCSI optimization, which can turn on flow control again on reboot, use the `no iscsi enable` command and save the configuration.

When you enable iSCSI on the switch, the following actions occur:

- Link-level flow control is globally enabled, if it is not already enabled, and PFC is disabled.
- iSCSI session snooping is enabled.
- iSCSI LLDP monitoring starts to automatically detect EqualLogic arrays.

The following message displays when you enable iSCSI on a switch and describes the configuration changes that are automatically performed:

```
%STKUNIT0-M:CP %IFMGR-5-IFM_ISCSI_ENABLE: iSCSI has been enabled causing flow
control to be
enabled on all interfaces. EQL detection and enabling iscsi profile-compellent
on an interface
may cause some automatic configurations to occur like jumbo frames on all ports
and no storm
control and spanning tree port-fast on the port of detection.
```

You can reconfigure any of the auto-provisioned configuration settings that result when you enable iSCSI on a switch.

When you disable the iSCSI feature, iSCSI resources are released and the detection of EqualLogic arrays using LLDP is disabled. Disabling iSCSI does not remove the MTU, flow control, portfast, or storm control configuration applied as a result of enabling iSCSI.

NOTE: By default, CAM allocation for iSCSI is set to 0. This disables session monitoring.

Default iSCSI Optimization Values

The following table lists the default values for the iSCSI optimization feature.

Table 30. iSCSI Optimization Defaults

Parameter	Default Value
iSCSI Optimization global setting	
iSCSI CoS mode (802.1p priority queue mapping)	
iSCSI CoS Packet classification	When you enable iSCSI, iSCSI packets are queued based on dot1p, instead of DSCP values.
VLAN priority tag	iSCSI flows are assigned by default to dot1p priority 4 without the `remark` setting.
DSCP	None: user-configurable.
Remark	Not configured.
iSCSI session aging time	10 minutes
iSCSI optimization target ports	iSCSI well-known ports 3260 and 860 are configured as default (with no IP address or name) but can be removed as any other configured target.

Parameter	Default Value
iSCSI session monitoring	Disabled. The CAM allocation for iSCSI is set to zero (0).

iSCSI Optimization Prerequisites

The following are iSCSI optimization prerequisites.

- iSCSI optimization requires LLDP on the switch. LLDP is enabled by default (refer to Link Layer Discovery Protocol (LLDP)).
- iSCSI optimization requires configuring two ingress ACL groups The ACL groups are allocated after iSCSI Optimization is configured.

Configuring iSCSI Optimization

To configure iSCSI optimization, use the following commands.

1. **For a non-DCB environment:** Enable session monitoring.

   CONFIGURATION mode

   ```
   cam-acl l2acl 4 ipv4acl 4 ipv6acl 0 ipv4qos 2 l2qos 1 l2pt 0 ipmacacl 0
   vman-qos 0 ecfmacl 0 fcoeacl 0 iscsioptacl 2
   ```

   NOTE: Content addressable memory (CAM) allocation is optional. If CAM is not allocated, the following features are disabled:

   - session monitoring
   - aging
   - class of service

   You can enable iSCSI even when allocated with zero (0) CAM blocks. However, if no CAM blocks are allocated, session monitoring is disabled and this information displays in the `show iscsi` command.

2. **For a non-DCB environment:** Enable iSCSI.

   CONFIGURATION mode

   ```
   iscsi enable
   ```

3. **For a DCB environment:** Configure iSCSI Optimization.

   EXEC Privilege mode

   iSCSI configuration: `copy CONFIG_TEMPLATE/iSCSI_DCB_Config running-config`.

   The configuration files are stored in the flash memory in the CONFIG_TEMPLATE file.

   NOTE: DCB/DCBx is enabled when you apply the iSCSI configuration in step 3. If you manually apply the iSCSI configuration by following steps 1 and 2, enable link layer discovery protocol (LLDP) before enabling iSCSI in step 2. You cannot disable LLDP if you enable iSCSI.

4. Save the configuration on the switch.

   EXEC Privilege mode

   ```
   write memory
   ```

5. Reload the switch.

   EXEC Privilege mode

   ```
   reload
   ```

   After the switch is reloaded, DCB/ DCBx and iSCSI monitoring are enabled.

6. (Optional) Configure the iSCSI target ports and optionally the IP addresses on which iSCSI communication is monitored.

   CONFIGURATION mode

   ```
   [no] iscsi target port tcp-port-1 [tcp-port-2...tcp-port-16] [ip-address
   address]
   ```

   - `tcp-port-n` is the TCP port number or a list of TCP port numbers on which the iSCSI target listens to requests. You can configure up to 16 target TCP ports on the switch in one command or multiple commands. The default is **860**, **3260**.

     Separate port numbers with a comma. If multiple IP addresses are mapped to a single TCP port, use the `no iscsi target port tcp-port-n` command to remove all IP addresses assigned to the TCP number.

     To delete a specific IP address from the TCP port, use the `no iscsi target port tcp-port-n ip-address address` command to specify the address to be deleted.
   - `ip-address` specifies the IP address of the iSCSI target. When you enter the `no` form of the command, and the TCP port you want to delete is one bound to a specific IP address, include the IP address value in the command.

     If multiple IP addresses are mapped to a single TCP port, use the `no iscsi target port` command to remove all IP addresses assigned to the TCP port number.

     To remove a single IP address from the TCP port, use the `no iscsi target port ip-address` command.

7. (Optional) Set the QoS policy that is applied to the iSCSI flows.

   CONFIGURATION mode

   ```
   [no] iscsi cos {enable | disable | dot1p vlan-priority-value [remark] | dscp
   dscp-value [remark]}
   ```

   - `enable`: enables the application of preferential QoS treatment to iSCSI traffic so that iSCSI packets are scheduled in the switch with a dot1p priority 4 regardless of the VLAN priority tag in the packet. The default is: iSCSI packets are handled with dotp1 priority 4 without `remark`.
   - `disable`: disables the application of preferential QoS treatment to iSCSI frames.
   - `dot1p vlan-priority-value`: specifies the virtual local area network (VLAN) priority tag assigned to incoming packets in an iSCSI session. The range is from 0 to 7. The default is: the dot1p value in ingress iSCSI frames is not changed and the same priority is used in iSCSI TLV advertisements if you do not enter the `iscsi priority-bits` command (Step 10).
   - `dscp dscp-value`: specifies the DSCP value assigned to incoming packets in an iSCSI session. The range is from 0 to 63. The default is: the DSCP value in ingress packets is not changed.
   - `remark`: marks incoming iSCSI packets with the configured dot1p or DSCP value when they egress the switch. The default is: the dot1 and DSCP values in egress packets are not changed.

8. (Optional) Set the aging time for iSCSI session monitoring.

   CONFIGURATION mode

   ```
   [no] iscsi aging time time.
   ```

The range is from 5 to 43,200 minutes.

The default is **10 minutes**.

9. (Optional) Configures DCBX to send iSCSI TLV advertisements.
   LLDP CONFIGURATION mode or INTERFACE LLDP CONFIGURATION mode

   `[no] advertise dcbx-app-tlv iscsi.`

   You can send iSCSI TLVs either globally or on a specified interface. The interface configuration takes priority over global configuration.

   The default is **Enabled**.
10. (Optional) Configures the advertised priority bitmap in iSCSI application TLVs.
    LLDP CONFIGURATION mode

    `[no] iscsi priority-bits.`

    The default is **4** (0x10 in the bitmap).
11. (Optional) Configures the auto-detection of Compellent arrays on a port.
    INTERFACE mode

    `[no] iscsi profile-compellent.`

    The default is: Compellent disk arrays are not detected.

Displaying iSCSI Optimization Information

To display information on iSCSI optimization, use the following `show` commands.

- Display the currently configured iSCSI settings.
  `show iscsi`
- Display information on active iSCSI sessions on the switch.
  `show iscsi sessions`
- Display detailed information on active iSCSI sessions on the switch . To display detailed information on specified iSCSI session, enter the session's iSCSI ID.
  `show iscsi sessions detailed [session` *`isid`*`]`
- Display all globally configured non-default iSCSI settings in the current Dell Networking OS session.
  `show run iscsi`

Examples of the `show iscsi` Commands

The following example shows the `show iscsi` command.

```
Dell#show iscsi
iSCSI is enabled
iSCSI session monitoring is disabled
iSCSI COS : dot1p is 4 no-remark
Session aging time: 10
Maximum number of connections is 256
------------------------------------------------
iSCSI Targets and TCP Ports:
------------------------------------------------
TCP Port Target IP Address
```

```
3260
860
```

The following example shows the `show iscsi session` command.

```
VLT PEER1

Dell#show iscsi session
Session 0:
-------------------------------------------------------------------------------------
Target: iqn.2001-05.com.equallogic:0-8a0906-0e70c2002-10a0018426a48c94-iom010
Initiator:
iqn.1991-05.com.microsoft:win-x9l8v27yajg
ISID: 400001370000
VLT PEER2

Session 0:
-------------------------------------------------------------------------------------
----
Target: iqn.2001-05.com.equallogic:0-8a0906-0f60c2002-0360018428d48c94-iom011
iqn.1991-05.com.microsoft:win-x9l8v27yajg
ISID: 400001370000
```

The following example shows the `show iscsi session detailed` command.

```
VLT PEER1

Dell# show iscsi session detailed
Session 0:
------------------------------------------------------------
Target:iqn.2010-11.com.ixia:ixload:iscsi-TG1
Initiator:iqn.2010-11.com.ixia.ixload:initiator-iscsi-2c
Up Time:00:00:01:28(DD:HH:MM:SS)
Time for aging out:00:00:09:34(DD:HH:MM:SS)
ISID:806978696102
Initiator Initiator Target Target Connection
IP Address TCP Port IP Address TCPPort ID
10.10.0.44 33345 10.10.0.101 3260 0

VLT PEER2
Session 0:
-------------------------------------------------------------
Target:iqn.2010-11.com.ixia:ixload:iscsi-TG1
Initiator:iqn.2010-11.com.ixia.ixload:initiator-iscsi-2c
Up Time:00:00:01:28(DD:HH:MM:SS)
Time for aging out:00:00:09:34(DD:HH:MM:SS)
ISID:806978696102
Initiator   Initiator  Target       Target   Connection
IP Address  TCP Port   IP Address   TCPPort  ID
10.10.0.53  33432      10.10.0.101  3260     0
```

23

Intermediate System to Intermediate System

Intermediate system to intermediate system (Is-IS) is supported on the S6000 platform.

-
- IS-IS is supported on the S6000 with Dell Networking OS 9.0(2.0).
- The IS-IS protocol is an interior gateway protocol (IGP) that uses a shortest-path-first algorithm. Dell Networking supports both IPv4 and IPv6 versions of IS-IS.
- The IS-IS protocol standards are listed in the Standards Compliance chapter.

IS-IS Protocol Overview

The IS-IS protocol, developed by the International Organization for Standardization (ISO), is an interior gateway protocol (IGP) that uses a shortest-path-first algorithm.

NOTE: This protocol supports routers passing both IP and OSI traffic, though the Dell Networking implementation supports only IP traffic.

IS-IS is organized hierarchically into routing domains and each router or system resides in at least one area. In IS-IS, routers are designated as Level 1, Level 2 or Level 1-2 systems. Level 1 routers only route traffic within an area, while Level 2 routers route traffic between areas. At its most basic, Level 1 systems route traffic within the area and any traffic destined for outside the area is sent to a Level 1-2 system. Level 2 systems manage destination paths for external routers. Only Level 2 routers can exchange data packets or routing information directly with external routers located outside of the routing domains. Level 1-2 systems manage both inter-area and intra-area traffic by maintaining two separate link databases; one for Level 1 routes and one for Level 2 routes. A Level 1-2 router does not advertise Level 2 routes to a Level 1 router.

To establish adjacencies, each IS-IS router sends different protocol data units (PDU). For IP traffic, the IP addressing information is included in the IS-IS hello PDUs and the link state PDUs (LSPs).

This brief overview is not intended to provide a complete understanding of IS-IS; for that, consult the documents listed in Multi-Topology IS-IS.

IS-IS Addressing

IS-IS PDUs require ISO-style addressing called network entity title (NET).

For those familiar with name-to-network service mapping point (NSAP) addresses, the composition of the NET is identical to an NSAP address, except the last byte is always 0. The NET is composed of the IS-IS area address, system ID, and N-selector. The last byte is the N-selector. All routers within an area have the same area portion. Level 1 routers route based on the system address portion of the address, while the Level 2 routers route based on the area address.

The NET length is variable, with a maximum of 20 bytes and a minimum of 8 bytes. It is composed of the following:

- **area address** — within your routing domain or area, each area must have a unique area value. The first byte is called the authority and format indicator (AFI).
- **system address** — the router's MAC address.
- **N-selector** — this is always 0.

The following illustration is an example of the ISO-style address to show the address format IS-IS uses. In this example, the first five bytes (47.0005.0001) are the area address. The system portion is 000c.000a. 4321 and the last byte is always 0.

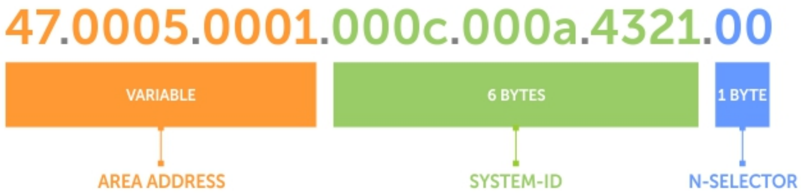


Figure 51. ISO Address Format

Multi-Topology IS-IS

Multi-topology IS-IS (MT IS-IS) allows you to create multiple IS-IS topologies on a single router with separate databases. Use this feature to place a virtual physical topology into logical routing domains, which can each support different routing and security policies.

All routers on a LAN or point-to-point must have at least one common supported topology when operating in Multi-Topology IS-IS mode. If IPv4 is the common supported topology between those two routers, adjacency can be formed. All topologies must share the same set of L1-L2 boundaries.

You must implement a wide metric-style globally on the autonomous system (AS) to run multi-topology IS-IS for IPv6 because the Type, Length, Value (TLVs) used to advertise IPv6 information in link-state packets (LSPs) are defined to use only extended metrics.

The multi-topology ID is shown in the first octet of the IS-IS packet. Certain MT topologies are assigned to serve predetermined purposes:

- MT ID #0: Equivalent to the “standard” topology.
- MT ID #1: Reserved for IPv4 in-band management purposes.
- MT ID #2: Reserved for IPv6 routing topology.
- MT ID #3: Reserved for IPv4 multicast routing topology.
- MT ID #4: Reserved for IPv6 multicast routing topology.
- MT ID #5: Reserved for IPv6 in-band management purposes.

Transition Mode

All routers in the area or domain must use the same type of IPv6 support, either single-topology or multi-topology. A router operating in multi-topology mode does not recognize the ability of the single-topology mode router to support IPv6 traffic, which leads to holes in the IPv6 topology.

While in Transition mode, both types of TLVs (single-topology and multi-topology) are sent in LSPs for all configured IPv6 addresses, but the router continues to operate in single-topology mode (that is, the topological restrictions of the single-topology mode remain in effect). Transition mode stops after all routers in the area or domain have been upgraded to support multi-topology IPv6. After all routers in the area or domain are operating in multi-topology IPv6 mode, the topological restrictions of single-topology mode are no longer in effect.

Interface Support

MT IS-IS is supported on physical Ethernet interfaces, physical synchronous optical network technologies (SONET) interfaces, port-channel interfaces (static and dynamic using LACP), and virtual local area network (VLAN) interfaces.

Adjacencies

Adjacencies on point-to-point interfaces are formed as usual, where IS-IS routers do not implement MT extensions.

If a local router does not participate in certain MTs, it does not advertise those MT IDs in its IS-IS hellos (IIHs) and so does not include that neighbor within its LSPs. If an MT ID is not detected in the remote side's IIHs, the local router does not include that neighbor within its LSPs. The local router does not form an adjacency if both routers do not have at least one common MT over the interface.

Graceful Restart

Graceful restart is supported on the platform for both Helper and Restart modes.

Graceful restart is a protocol-based mechanism that preserves the forwarding table of the restarting router and its neighbors for a specified period to minimize the loss of packets. A graceful-restart router does not immediately assume that a neighbor is permanently down and does not trigger a topology change.

Normally, when an IS-IS router is restarted, temporary disruption of routing occurs due to events in both the restarting router and the neighbors of the restarting router. When a router goes down without a graceful restart, there is a potential to lose access to parts of the network due to the necessity of network topology changes.

IS-IS graceful restart recognizes that in a modern router, the control plane and data plane are functionally separate. Restarting the control plane functionality (such as the failover of the active route processor module (RPM) to the backup in a redundant configuration) should not necessarily interrupt data packet forwarding. This behavior is supported because the forwarding tables previously computed by an active RPM have been downloaded into the forwarding information base (FIB) on the line cards (the data plane). For packets that have existing FIB/content addressable memory (CAM) entries, forwarding between ingress and egress ports can continue uninterrupted while the control plane IS-IS process comes back to full functionality and rebuilds its routing tables.

A new TLV (the Restart TLV) is introduced in the IIH PDUs, indicating that the router supports graceful restart.

Timers

Three timers are used to support IS-IS graceful restart functionality. After you enable graceful restart, these timers manage the graceful restart process.

There are three times, T1, T2, and T3.

- The T1 timer specifies the wait time before unacknowledged restart requests are generated. This is the interval before the system sends a Restart Request (an IIH with the RR bit set in Restart TLV) until the complete sequence number PDU (CSNP) is received from the helping router. You can set the duration to a specific amount of time (seconds) or a number of attempts.
- The T2 timer is the maximum time that the system waits for LSP database synchronization. This timer applies to the database type (level-1, level-2, or both).
- The T3 timer sets the overall wait time after which the router determines that it has failed to achieve database synchronization (by setting the overload bit in its own LSP). You can base this timer on adjacency settings with the value derived from adjacent routers that are engaged in graceful restart recovery (the minimum of all the Remaining Time values advertised by the neighbors) or by setting a specific amount of time manually.

Implementation Information

IS-IS implementation supports one instance of IS-IS and six areas.

You can configure the system as a Level 1 router, a Level 2 router, or a Level 1-2 router. For IPv6, the IPv4 implementation has been expanded to include two new type, length, values (TLVs) in the PDU that carry information required for IPv6 routing. The new TLVs are *IPv6 Reachability* and *IPv6 Interface Address*. Also, a new IPv6 protocol identifier has also been included in the supported TLVs. The new TLVs use the extended metrics and up/down bit semantics.

Multi-topology IS-IS adds TLVs:

- **MT TLV** — contains one or more Multi-Topology IDs in which the router participates. This TLV is included in IIH and the first fragment of an LSP.
- **MT Intermediate Systems TLV** — appears for every topology a node supports. An MT ID is added to the extended IS reachability TLV type 22.
- **MT Reachable IPv4 Prefixes TLV** — appears for each IPv4 an IS announces for a given MT ID. Its structure is aligned with the extended IS Reachability TLV Type 236 and it adds an MT ID.
- **MT Reachable IPv6 Prefixes TLV** — appears for each IPv6 an IS announces for a given MT ID. Its structure is aligned with the extended IS Reachability TLV Type 236 and add an MT ID.

By default, Dell Networking OS supports dynamic host name exchange to assist with troubleshooting and configuration. By assigning a name to an IS-IS NET address, you can track IS-IS information on that address easier. Dell Networking OS does not support ISO CLNS routing; however, the ISO NET format is supported for addressing.

To support IPv6, the Dell Networking implementation of IS-IS performs the following tasks:

- Advertises IPv6 information in the PDUs.
- Processes IPv6 information received in the PDUs.
- Computes routes to IPv6 destinations.
- Downloads IPv6 routes to the RTM for installing in the FIB.

- Accepts external IPv6 information and advertises this information in the PDUs.

The following table lists the default IS-IS values.

Table 31. IS-IS Default Values

IS-IS Parameter	Default Value
Complete sequence number PDU (CSNP) interval	10 seconds
IS-to-IS hello PDU interval	10 seconds
IS-IS interface metric	10
Metric style	Narrow
Designated Router priority	64
Circuit Type	Level 1 and Level 2
IS Type	Level 1 and Level 2
Equal Cost Multi Paths	16

Configuration Information

To use IS-IS, you must configure and enable IS-IS in two or three modes: CONFIGURATION ROUTER ISIS, CONFIGURATION INTERFACE, and (when configuring for IPv6) ADDRESS-FAMILY mode. Commands in ROUTER ISIS mode configure IS-IS globally, while commands executed in INTERFACE mode enable and configure IS-IS features on that interface only. Commands in the ADDRESS-FAMILY mode are specific to IPv6.

NOTE: When using the IS-IS routing protocol to exchange IPv6 routing information and to determine destination reachability, you can route IPv6 along with IPv4 while using a single intra-domain routing protocol. The configuration commands allow you to enable and disable IPv6 routing and to configure or remove IPv6 prefixes on links.

Except where identified, the commands described in this chapter apply to both IPv4 and IPv6 versions of IS-IS.

Configuration Tasks for IS-IS

The following describes the configuration tasks for IS-IS.

- Enabling IS-IS
- Configure Multi-Topology IS-IS (MT IS-IS)
- Configuring IS-IS Graceful Restart
- Changing LSP Attributes
- Configuring the IS-IS Metric Style
- Configuring IS-IS Cost
- Changing the IS-Type
- Controlling Routing Updates
- Configuring Authentication Passwords
- Setting the Overload Bit
- Debuging IS-IS

Enabling IS-IS

By default, IS-IS is not enabled.
The system supports one instance of IS-IS. To enable IS-IS globally, create an IS-IS routing process and assign a NET address. To exchange protocol information with neighbors, enable IS-IS on an interface, instead of on a network as with other routing protocols.

In IS-IS, neighbors form adjacencies only when they are same IS type. For example, a Level 1 router never forms an adjacency with a Level 2 router. A Level 1-2 router forms Level 1 adjacencies with a neighboring Level 1 router and forms Level 2 adjacencies with a neighboring Level 2 router.

NOTE: Even though you enable IS-IS globally, enable the IS-IS process on an interface for the IS-IS process to exchange protocol information and form adjacencies.

To configure IS-IS globally, use the following commands.

1. Create an IS-IS routing process.
   CONFIGURATION mode

   ```
   router isis [tag]
   ```

   tag: (optional) identifies the name of the IS-IS process.
2. Configure an IS-IS network entity title (NET) for a routing process.
   ROUTER ISIS mode

   ```
   net network-entity-title
   ```

   Specify the area address and system ID for an IS-IS routing process. The last byte must be 00.

   For more information about configuring a NET, refer to IS-IS Addressing.
3. Enter the interface configuration mode.
   CONFIGURATION mode

   ```
   interface interface
   ```

   Enter the keyword `interface` then the type of interface and slot/port information:

   - For a 1-Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
   - For the Loopback interface on the RPM, enter the keyword `loopback` then a number from 0 to 16383.
   - For a port channel, enter the keywords `port-channel` then a number.
   - For a SONET interface, enter the keyword `sonet` then the slot/port information.
   - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
   - For a VLAN, enter the keyword `vlan` then a number from 1 to 4094.
4. Enter an IPv4 Address.
   INTERFACE mode

   ```
   ip address ip-address mask
   ```

   Assign an IP address and mask to the interface.

The IP address must be on the same subnet as other IS-IS neighbors, but the IP address does not need to relate to the NET address.

5. Enter an IPv6 Address.

   INTERFACE mode

   ```
   ipv6 address ipv6-address mask
   ```

   - *`ipv6 address`*: x:x:x:x::x
   - *`mask`*: The prefix length is from 0 to 128.

   The IPv6 address must be on the same subnet as other IS-IS neighbors, but the IP address does not need to relate to the NET address.

6. Enable IS-IS on the IPv4 interface.

   ROUTER ISIS mode

   ```
   ip router isis [tag]
   ```

   If you configure a tag variable, it must be the same as the *`tag`* variable assigned in step 1.

7. Enable IS-IS on the IPv6 interface.

   ROUTER ISIS mode

   ```
   ipv6 router isis [tag]
   ```

   If you configure a tag variable, it must be the same as the *`tag`* variable assigned in step 1.

Examples of the `show isis` Commands

The default IS type is **level-1-2**. To change the IS type to Level 1 only or Level 2 only, use the `is-type` command in ROUTER ISIS mode.

To view the IS-IS configuration, enter the `show isis protocol` command in EXEC Privilege mode or the `show config` command in ROUTER ISIS mode.

```
Dell#show isis protocol
IS-IS Router: <Null Tag>
System Id: EEEE.EEEE.EEEE IS-Type: level-1-2
Manual area address(es):
  47.0004.004d.0001
Routing for area address(es):
  21.2223.2425.2627.2829.3031.3233
  47.0004.004d.0001
Interfaces supported by IS-IS:
  Vlan 2
  GigabitEthernet 4/22
  Loopback 0
Redistributing:
Distance: 115
Generate narrow metrics: level-1-2
Accept narrow metrics:   level-1-2
Generate wide metrics:   none
Accept wide metrics:     none
Dell#
```

To view IS-IS protocol statistics, use the `show isis traffic` command in EXEC Privilege mode.

```
Dell#show isis traffic
  IS-IS: Level-1 Hellos (sent/rcvd) : 4272/1538
```

```
  IS-IS: Level-2 Hellos (sent/rcvd) : 4272/1538
  IS-IS: PTP Hellos (sent/rcvd) : 0/0
  IS-IS: Level-1 LSPs sourced (new/refresh) : 0/0
  IS-IS: Level-2 LSPs sourced (new/refresh) : 0/0
  IS-IS: Level-1 LSPs flooded (sent/rcvd) : 32/19
  IS-IS: Level-2 LSPs flooded (sent/rcvd) : 32/17
  IS-IS: Level-1 LSPs CSNPs (sent/rcvd) : 1538/0
  IS-IS: Level-2 LSPs CSNPs (sent/rcvd) : 1534/0
  IS-IS: Level-1 LSPs PSNPs (sent/rcvd) : 0/0
  IS-IS: Level-2 LSPs PSNPs (sent/rcvd) : 0/0
  IS-IS: Level-1 DR Elections : 2
  IS-IS: Level-2 DR Elections : 2
  IS-IS: Level-1 SPF Calculations : 29
  IS-IS: Level-2 SPF Calculations : 29
  IS-IS: LSP checksum errors received : 0
  IS-IS: LSP authentication failures : 0
Dell#
```

You can assign more NET addresses, but the System ID portion of the NET address must remain the same. Dell Networking OS supports up to six area addresses.

Some address considerations are:

- In order to be neighbors, configure Level 1 routers with at least one common area address.
- A Level 2 router becomes a neighbor with another Level 2 router regardless of the area address configured. However, if the area addresses are different, the link between the Level 2 routers is only at Level 2.

Configuring Multi-Topology IS-IS (MT IS-IS)

To configure multi-topology IS-IS (MT IS-IS), use the following commands.

1. Enable multi-topology IS-IS for IPv6.

   ROUTER ISIS AF IPV6 mode

   ```
   multi-topology [transition]
   ```

   Enter the keyword *`transition`* to allow an IS-IS IPv6 user to continue to use single-topology mode while upgrading to multi-topology mode. After every router has been configured with the *`transition`* keyword, and all the routers are in MT IS-IS IPv6 mode, you can remove the *`transition`* keyword on each router.

   NOTE: When you do not enable transition mode, you do not have IPv6 connectivity between routers operating in single-topology mode and routers operating in multi-topology mode.

2. Exclude this router from other router’s SPF calculations.

   ROUTER ISIS AF IPV6 mode

   ```
   set-overload-bit
   ```

3. Set the minimum interval between SPF calculations.

   ROUTER ISIS AF IPV6 mode

   ```
   spf-interval [level-l | level-2 | interval] [initial_wait_interval
   [second_wait_interval]]
   ```

Use this command for IPv6 route computation only when you enable multi-topology. If using single-topology mode, to apply to both IPv4 and IPv6 route computations, use the `spf-interval` command in CONFIG ROUTER ISIS mode.

4. Implement a *wide metric-style* globally.

   ROUTER ISIS AF IPV6 mode

   ```
   isis ipv6 metric metric-value [level-1 | level-2 | level-1-2]
   ```

   To configure wide or wide transition metric style, the cost can be between 0 and 16,777,215.

Configuring IS-IS Graceful Restart

To enable IS-IS graceful restart globally, use the following commands. Additionally, you can implement optional commands to enable the graceful restart settings.

- Enable graceful restart on ISIS processes.

  ROUTER-ISIS mode

  ```
  graceful-restart ietf
  ```
- Configure the time during which the graceful restart attempt is prevented.

  ROUTER-ISIS mode

  ```
  graceful-restart interval minutes
  ```

  The range is from 1 to 120 minutes.

  The default is **5 minutes**.
- Enable the graceful restart maximum wait time before a restarting peer comes up.

  ROUTER-ISIS mode

  ```
  graceful-restart restart-wait seconds
  ```

  When implementing this command, be sure to set the t3 timer to adjacency on the restarting router.

  The range is from 1 to 120 minutes.

  The default is **30 seconds**.
- Configure the time that the graceful restart timer T1 defines for a restarting router to use for each interface, as an interval before regenerating Restart Request (an IIH with RR bit set in Restart TLV) after waiting for an acknowledgement.

  ROUTER-ISIS mode

  ```
  graceful-restart t1 {interval seconds | retry-times value}
  ```

  - `interval`: wait time (the range is from 5 to 120. The default is **5**.)
  - `retry-times`: number of times an unacknowledged restart request is sent before the restarting router gives up the graceful restart engagement with the neighbor. (The range is from 1 to 10 attempts. The default is **1**.)
- Configure the time for the graceful restart timer T2 that a restarting router uses as the wait time for each database to synchronize.

  ROUTER-ISIS mode

  ```
  graceful-restart t2 {level-1 | level-2} seconds
  ```

  - `level-1`, `level-2`: identifies the database instance type to which the wait interval applies.

  The range is from 5 to 120 seconds.

  The default is **30 seconds**.

- Configure graceful restart timer T3 to set the time used by the restarting router as an overall maximum time to wait for database synchronization to complete.

  ROUTER-ISIS mode

  `graceful-restart t3 {adjacency | manual *seconds*}`

  - `adjacency`: the restarting router receives the remaining time value from its peer and adjusts its T3 value so if user has configured this option.
  - `manual`: allows you to specify a fixed value that the restarting router should use.

  The range is from 50 to 120 seconds.

  The default is **30 seconds**.

Examples of the `show isis` Commands

NOTE: If this timer expires before the synchronization has completed, the restarting router sends the overload bit in the LSP. The 'overload' bit is an indication to the receiving router that database synchronization did not complete at the restarting router.

To view all graceful restart-related configurations, use the `show isis graceful-restart detail` command in EXEC Privilege mode.

```
Dell#show isis graceful-restart detail
Configured Timer Value
======================
Graceful Restart       : Enabled
Interval/Blackout time : 1 min
T3 Timer               : Manual
T3 Timeout Value       : 30
T2 Timeout Value       : 30 (level-1), 30 (level-2)
T1 Timeout Value       : 5, retry count: 1
Adjacency wait time    : 30

Operational Timer Value
======================
Current Mode/State     : Normal/RUNNING
T3 Time left           : 0
T2 Time left           : 0 (level-1), 0 (level-2)
Restart ACK rcv count  : 0 (level-1), 0 (level-2)
Restart Req rcv count  : 0 (level-1), 0 (level-2)
Suppress Adj rcv count : 0 (level-1), 0 (level-2)
Restart CSNP rcv count : 0 (level-1), 0 (level-2)
Database Sync count    : 0 (level-1), 0 (level-2)

Circuit GigabitEthernet 2/10:
  Mode: Normal L1-State:NORMAL, L2-State: NORMAL

  L1: Send/Receive: RR:0/0, RA: 0/0, SA:0/0
    T1 time left: 0, retry count left:0

  L2: Send/Receive: RR:0/0, RA: 0/0, SA:0/0
    T1 time left: 0, retry count left:0
Dell#
```

To view all interfaces configured with IS-IS routing along with the defaults, use the `show isis interface` command in EXEC Privilege mode.

```
Dell#show isis interface G1/34
GigabitEthernet 2/10 is up, line protocol is up
  MTU 1497, Encapsulation SAP
  Routing Protocol: IS-IS
    Circuit Type: Level-1-2
    Interface Index 0x62cc03a, Local circuit ID 1
    Level-1 Metric: 10, Priority: 64, Circuit ID: 0000.0000.000B.01
        Hello Interval: 10, Hello Multiplier: 3, CSNP Interval: 10
    Number of active level-1 adjacencies: 1
    Level-2 Metric: 10, Priority: 64, Circuit ID: 0000.0000.000B.01
        Hello Interval: 10, Hello Multiplier: 3, CSNP Interval: 10
    Number of active level-2 adjacencies: 1
    Next IS-IS LAN Level-1 Hello in 4 seconds
    Next IS-IS LAN Level-2 Hello in 6 seconds
    LSP Interval: 33 Next IS-IS LAN Level-1 Hello in 4 seconds
    Next IS-IS LAN Level-2 Hello in 6 seconds
    LSP Interval: 33
Restart Capable Neighbors: 2, In Start: 0, In Restart: 0
Dell#
```

Changing LSP Attributes

IS-IS routers flood link state PDUs (LSPs) to exchange routing information. LSP attributes include the generation interval, maximum transmission unit (MTU) or size, and the refresh interval.
You can modify the LSP attribute defaults, but it is not necessary.

To change the defaults, use any or all of the following commands.

- Set interval between LSP generation.

  ROUTER ISIS mode

  `lsp-gen-interval [level-1 | level-2] seconds`

  – *seconds*: the range is from 0 to 120.

  The default is **5 seconds**.

  The default level is **Level 1**.
- Set the LSP size.

  ROUTER ISIS mode

  `lsp-mtu size`

  – *size*: the range is from 128 to 9195.

  The default is **1497**.
- Set the LSP refresh interval.

  ROUTER ISIS mode

  `lsp-refresh-interval seconds`

  – *seconds*: the range is from 1 to 65535.

  The default is **900 seconds**.
- Set the maximum time LSPs lifetime.

  ROUTER ISIS mode

`max-lsp-lifetime` *`seconds`*

– *`seconds`*: the range is from 1 to 65535.

The default is **1200 seconds**.

Example of Viewing IS-IS Configuration (ROUTER ISIS Mode)

To view the configuration, use the `show config` command in ROUTER ISIS mode or the `show running-config isis` command in EXEC Privilege mode.

```
Dell#show running-config isis
!
router isis
  lsp-refresh-interval 902
  net 47.0005.0001.000C.000A.4321.00
  net 51.0005.0001.000C.000A.4321.00
Dell#
```

Configuring the IS-IS Metric Style

All IS-IS links or interfaces are associated with a cost that is used in the shortest path first (SPF) calculations. The possible cost varies depending on the metric style supported.

If you configure narrow, transition, or narrow transition metric style, the cost can be a number between 0 and 63. If you configure wide or wide transition metric style, the cost can be a number between 0 and 16,777,215. Dell Networking OS supports five different metric styles: narrow, wide, transition, narrow transition, and wide transition.

By default, Dell Networking OS generates and receives narrow metric values. Matrixes or costs higher than 63 are not supported. To accept or generate routes with a higher metric, you must change the metric style of the IS-IS process. For example, if you configure the metric as narrow, and a link state PDU (LSP) with wide metrics is received, the route is not installed.

Dell Networking OS supports the following IS-IS metric styles.

Table 32. Metric Styles

Metric Style	Characteristics	Cost Range Supported on IS-IS Interfaces
narrow	Sends and accepts narrow or old TLVs (Type, Length, Value).	0 to 63
wide	Sends and accepts wide or new TLVs.	0 to 16777215
transition	Sends both wide (new) and narrow (old) TLVs.	0 to 63
narrow transition	Sends narrow (old) TLVs and accepts both narrow (old) and wide (new) TLVs.	0 to 63
wide transition	Sends wide (new) TLVs and accepts both narrow (old) and wide (new) TLVs.	0 to 16777215

To change the IS-IS metric style of the IS-IS process, use the following command.

- Set the metric style for the IS-IS process.

  ROUTER ISIS mode

  ```
  metric-style {narrow [transition] | transition | wide [transition]} [level-1
  | level-2]
  ```

  The default is **narrow**.

  The default is Level 1 and Level 2 (**level-1–2**)

To view which metric types are generated and received, use the `show isis protocol` command in EXEC Privilege mode. The IS-IS matrixes settings are in bold.

Example of Viewing IS-IS Metric Types

```
Dell#show isis protocol
IS-IS Router: <Null Tag>
  System Id: EEEE.EEEE.EEEE IS-Type: level-1-2
  Manual area address(es):
    47.0004.004d.0001
  Routing for area address(es):
    21.2223.2425.2627.2829.3031.3233
    47.0004.004d.0001
  Interfaces supported by IS-IS:
    Vlan 2
    GigabitEthernet 4/22
    Loopback 0
  Redistributing:
  Distance: 115
  Generate narrow metrics: level-1-2
  Accept narrow metrics:   level-1-2
  Generate wide metrics:   none
  Accept wide metrics:     none
Dell#
```

Configuring the IS-IS Cost

When you change from one IS-IS metric style to another, the IS-IS metric value could be affected. For each interface with IS-IS enabled, you can assign a cost or metric that is used in the link state calculation. To change the metric or cost of the interface, use the following commands.

- Assign an IS-IS metric.

  INTERFACE mode

  `isis metric default-metric [level-1 | level-2]`

  - *`default-metric`*: the range is from 0 to 63 if the metric-style is narrow, narrow-transition, or transition.

  The range is from 0 to 16777215 if the metric style is wide or wide transition.
- Assign a metric for an IPv6 link or interface.

  INTERFACE mode

  `isis ipv6 metric default-metric [level-1 | level-2]`

  - *`default-metric`*: the range is from 0 to 63 for narrow and transition metric styles. The range is from 0 to 16777215 for wide metric styles.

  The default is **10**.

The default level is **level-1**.

For more information about this command, refer to Configuring the IS-IS Metric Style.

The following table describes the correct value range for the `isis metric` command.

Metric Sytle	Correct Value Range
wide	0 to 16777215
narrow	0 to 63
wide transition	0 to 16777215
narrow transition	0 to 63
transition	0 to 63

To view the interface's current metric, use the `show config` command in INTERFACE mode or the `show isis interface` command in EXEC Privilege mode.

Configuring the Distance of a Route

To configure the distance for a route, use the following command.

- Configure the distance for a route.
  ROUTER ISIS mode

  ```
  distance
  ```

Changing the IS-Type

To change the IS-type, use the following commands.
You can configure the system to act as a Level 1 router, a Level 1-2 router, or a Level 2 router.

To change the IS-type for the router, use the following commands.

- Configure IS-IS operating level for a router.
  ROUTER ISIS mode

  ```
  is-type {level-1 | level-1-2 | level-2-only}
  ```

  Default is **level-1-2**.
- Change the IS-type for the IS-IS process.
  ROUTER ISIS mode

  ```
  is-type {level-1 | level-1-2 | level-2}
  ```

Example of the `show isis database` Command to View Level 1-2 Link State Databases

To view which IS-type is configured, use the `show isis protocol` command in EXEC Privilege mode. The `show config` command in ROUTER ISIS mode displays only non-default information. If you do not change the IS-type, the default value (**level-1-2**) is not displayed.

The default is Level 1-2 router. When the IS-type is Level 1-2, the software maintains two Link State databases, one for each level. To view the Link State databases, use the `show isis database` command.

```
Dell#show isis database
IS-IS Level-1 Link State Database
LSPID           LSP Seq Num  LSP Checksum LSP Holdtime ATT/P/OL
B233.00-00      0x00000003   0x07BF       1088         0/0/0
eljefe.00-00 *  0x00000009   0xF76A       1126         0/0/0
eljefe.01-00 *  0x00000001   0x68DF       1122         0/0/0
eljefe.02-00 *  0x00000001   0x2E7F       1113         0/0/0
Force10.00-00   0x00000002   0xD1A7       1102         0/0/0
IS-IS Level-2 Link State Database
LSPID           LSP Seq Num  LSP Checksum LSP Holdtime ATT/P/OL
B233.00-00      0x00000006   0xC38A       1124         0/0/0
eljefe.00-00 *  0x0000000D   0x51C6       1129         0/0/0
eljefe.01-00 *  0x00000001   0x68DF       1122         0/0/0
eljefe.02-00 *  0x00000001   0x2E7F       1113         0/0/0
Force10.00-00   0x00000004   0xCDA9       1107         0/0/0

Dell#
```

Controlling Routing Updates

To control the source of IS-IS route information, use the following command.

- Disable a specific interface from sending or receiving IS-IS routing information.

  ROUTER ISIS mode

  ```
  passive-interface interface
  ```

  - For a 1-Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
  - For the Loopback interface on the RPM, enter the keyword `loopback` then a number from 0 to 16383.
  - For a port channel, enter the keywords `port-channel` then a number.
  - For a SONET interface, enter the keyword `sonet` then the slot/port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/ port information.
  - For a VLAN, enter the keyword `vlan` then a number from 1 to 4094.

Distribute Routes

Another method of controlling routing information is to filter the information through a prefix list.

Prefix lists are applied to incoming or outgoing routes and routes must meet the conditions of the prefix lists or Dell Networking OS does not install the route in the routing table. The prefix lists are globally applied on all interfaces running IS-IS.

Configure the prefix list in PREFIX LIST mode prior to assigning it to the IS-IS process. For configuration information on prefix lists, refer to Access Control Lists (ACLs).

Applying IPv4 Routes

To apply prefix lists to incoming or outgoing IPv4 routes, use the following commands.

NOTE: These commands apply to IPv4 IS-IS only. To apply prefix lists to IPv6 routes, use ADDRESS-FAMILY IPV6 mode, shown later.

- Apply a configured prefix list to all incoming IPv4 IS-IS routes.

  ROUTER ISIS mode

```
distribute-list prefix-list-name in [interface]
```

  - Enter the type of interface and slot/port information:
  - For a 1-Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
  - For the Loopback interface on the RPM, enter the keyword `loopback` then a number from 0 to 16383.
  - For a port channel, enter the keywords `port-channel` then a number.
  - For a SONET interface, enter the keyword `sonet` then the slot/port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
  - For a VLAN, enter the keyword `vlan` then a number from 1 to 4094.
- Apply a configured prefix list to all outgoing IPv4 IS-IS routes.

  ROUTER ISIS mode

```
distribute-list prefix-list-name out [bgp as-number | connected | ospf
process-id | rip | static]
```

  You can configure one of the optional parameters:

  - `connected`: for directly connected routes.
  - `ospf process-id`: for OSPF routes only.
  - `rip`: for RIP routes only.
  - `static`: for user-configured routes.
  - `bgp`: for BGP routes only.
- Deny RTM download for pre-existing redistributed IPv4 routes.

  ROUTER ISIS mode

```
distribute-list redistributed-override in
```

Applying IPv6 Routes

To apply prefix lists to incoming or outgoing IPv6 routes, use the following commands.

NOTE: These commands apply to IPv6 IS-IS only. To apply prefix lists to IPv4 routes, use ROUTER ISIS mode, previously shown.

- Apply a configured prefix list to all incoming IPv6 IS-IS routes.

  ROUTER ISIS-AF IPV6 mode

```
distribute-list prefix-list-name in [interface]
```

  Enter the type of interface and slot/port information:

  - For a 1-Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
  - For the Loopback interface on the RPM, enter the keyword `loopback` then a number from 0 to 16383.
  - For a port channel, enter the keywords `port-channel` then a number.
  - For a SONET interface, enter the keyword `sonet` then the slot/port information.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.

  - For a VLAN, enter the keyword `vlan` then a number from 1 to 4094.
- Apply a configured prefix list to all outgoing IPv6 IS-IS routes.

  ROUTER ISIS-AF IPV6 mode

  ```
  distribute-list prefix-list-name out [bgp as-number | connected | ospf
  process-id | rip | static]
  ```

  You can configure one of the optional parameters:

  - `connected`: for directly connected routes.
  - `ospf` *`process-id`*: for OSPF routes only.
  - `rip`: for RIP routes only.
  - `static`: for user-configured routes.
  - `bgp`: for BGP routes only.
- Deny RTM download for pre-existing redistributed IPv6 routes.

  ROUTER ISIS-AF IPV6 mode

  ```
  distribute-list redistributed-override in
  ```

Redistributing IPv4 Routes

In addition to filtering routes, you can add routes from other routing instances or protocols to the IS-IS process. With the `redistribute` command syntax, you can include BGP, OSPF, RIP, static, or directly connected routes in the IS-IS process.

NOTE: Do not route iBGP routes to IS-IS unless there are route-maps associated with the IS-IS redistribution.

To add routes from other routing instances or protocols, use the following commands.

NOTE: These commands apply to IPv4 IS-IS only. To apply prefix lists to IPv6 routes, use ADDRESS-FAMILY IPV6 mode, shown later.

- Include BGP, directly connected, RIP, or user-configured (static) routes in IS-IS.

  ROUTER ISIS mode

  ```
  redistribute {bgp as-number | connected | rip | static} [level-1 level-1-2 |
  level-2] [metric metric-value] [metric-type {external | internal}] [route-map
  map-name]
  ```

  Configure the following parameters:

  - `level-1`, `level-1-2`, or `level-2`: assign all redistributed routes to a level. The default is **level-2**.
  - *`metric-value`* the range is from 0 to 16777215. The default is **0**.
  - *`metric-type`*: choose either external or internal. The default is **internal**.
  - *`map-name`*: enter the name of a configured route map.
- Include specific OSPF routes in IS-IS.

  ROUTER ISIS mode

  ```
  redistribute ospf process-id [level-1| level-1-2 | level-2] [metric value]
  [match external {1 | 2} | match internal] [metric-type {external | internal}]
  [route-map map-name]
  ```

  Configure the following parameters:

  - *`process-id`* the range is from 1 to 65535.
  - `level-1`, `level-1-2`, or `level-2`: assign all redistributed routes to a level. The default is **level-2**.
  - `metric` *`value`* the range is from 0 to 16777215. The default is **0**.
  - `match external` the range is from 1 or 2.
  - `match internal`
  - `metric-type`: external or internal.
  - *`map-name`*: enter the name of a configured route map.

Redistributing IPv6 Routes

To add routes from other routing instances or protocols, use the following commands.

NOTE: These commands apply to IPv6 IS-IS only. To apply prefix lists to IPv4 routes, use the ROUTER ISIS mode previously shown.

- Include BGP, directly connected, RIP, or user-configured (static) routes in IS-IS.

  ROUTER ISIS mode

  ```
  redistribute {bgp as-number | connected | rip | static} [level-1 level-1-2 |
  level-2] [metric metric-value] [metric-type {external | internal}] [route-map
  map-name]
  ```

  Configure the following parameters:

  - `level-1`, `level-1-2`, or `level-2`: assign all redistributed routes to a level. The default is **level-2**.
  - *`metric-value`*: the range is from 0 to 16777215. The default is **0**.
  - *`metric-type`*: choose either external or internal. The default is **internal**.
  - *`map-name`*: enter the name of a configured route map.
- Include specific OSPF routes in IS-IS.ROUTER ISIS mode

  ```
  redistribute ospf process-id [level-1| level-1-2 | level-2] [metric value]
  [match external {1 | 2} | match internal] [metric-type {external | internal}]
  [route-map map-name]
  ```

  Configure the following parameters:

  - *`process-id`*: the range is from 1 to 65535.
  - `level-1`, `level-1-2`, or `level-2`: assign all redistributed routes to a level. The default is **level-2**.
  - `metric` *`value`*: the range is from 0 to 16777215. The default is **0**.
  - `metric` *`value`*: the range is from 0 to 16777215. The default is **0**.
  - `match external`: the range is 1 or 2.
  - `match internal`
  - `metric-type`: external or internal.
  - *`map-name`*: name of a configured route map.

To view the IS-IS configuration globally (including both IPv4 and IPv6 settings), use the `show running-config isis` command in EXEC Privilege mode. To view the current IPv4 IS-IS configuration, use the `show config` command in ROUTER ISIS mode. To view the current IPv6 IS-IS configuration, use the `show config` command in ROUTER ISIS-ADDRESS FAMILY IPV6 mode.

Configuring Authentication Passwords

You can assign an authentication password for routers in Level 1 and for routers in Level 2. Because Level 1 and Level 2 routers do not communicate with each other, you can assign different passwords for Level 1 routers and for Level 2 routers. However, if you want the routers in the level to communicate with each other, configure them with the same password.

To configure a simple text password, use the following commands.

- Configure authentication password for an area.

  ROUTER ISIS mode

  ```
  area-password [hmac-md5] password
  ```

  FTOS supports HMAC-MD5 authentication.

  This password is inserted in Level 1 LSPs, Complete SNPs, and Partial SNPs.
- Set the authentication password for a routing domain.

  ROUTER ISIS mode

  ```
  domain-password [encryption-type | hmac-md5] password
  ```

  FTOS supports both DES and HMAC-MD5 authentication methods.

  This password is inserted in Level 2 LSPs, Complete SNPs, and Partial SNPs.

To view the passwords, use the `show config` command in ROUTER ISIS mode or the `show running-config isis` command in EXEC Privilege mode.

To remove a password, use either the `no area-password` or `no domain-password` commands in ROUTER ISIS mode.

Setting the Overload Bit

Another use for the overload bit is to prevent other routers from using this router as an intermediate hop in their shortest path first (SPF) calculations. For example, if the IS-IS routing database is out of memory and cannot accept new LSPs, Dell Networking OS sets the overload bit and IS-IS traffic continues to transit the system.

To set or remove the overload bit manually, use the following commands.

- Set the overload bit in LSPs.

  ROUTER ISIS mode

  ```
  set-overload-bit
  ```

  This setting prevents other routers from using it as an intermediate hop in their shortest path first (SPF) calculations.
- Remove the overload bit.

  ROUTER ISIS mode

  ```
  no set-overload-bit
  ```

Example of Viewing the Overload Bit Setting

When the bit is set, a 1 is placed in the *OL* column in the `show isis database` command output. The overload bit is set in both the Level-1 and Level-2 database because the IS type for the router is Level-1-2.

```
Dell#show isis database
IS-IS Level-1 Link State Database
LSPID           LSP Seq Num LSP Checksum LSP Holdtime ATT/P/OL
B233.00-00     0x00000003   0x07BF       1074         0/0/0
eljefe.00-00 * 0x0000000A   0xF963       1196         0/0/1
eljefe.01-00 * 0x00000001   0x68DF       1108         0/0/0
eljefe.02-00 * 0x00000001   0x2E7F       1099         0/0/0
Force10.00-00  0x00000002   0xD1A7       1088         0/0/0
IS-IS Level-2 Link State Database
LSPID           LSP Seq Num LSP Checksum LSP Holdtime ATT/P/OL
B233.00-00     0x00000006   0xC38A       1110         0/0/0
eljefe.00-00 * 0x0000000E   0x53BF       1196         0/0/1
eljefe.01-00 * 0x00000001   0x68DF       1108         0/0/0
eljefe.02-00 * 0x00000001   0x2E7F       1099         0/0/0
Force10.00-00  0x00000004   0xCDA9       1093         0/0/0
Dell#
```

Debugging IS-IS

To debug IS-IS processes, use the following commands.

- View all IS-IS information.

  EXEC Privilege mode

  ```
  debug isis
  ```
- View information on all adjacency-related activity (for example, hello packets that are sent and received).

  EXEC Privilege mode

  ```
  debug isis adj-packets [interface]
  ```

  To view specific information, enter the following optional parameter:

  - *`interface`*: Enter the type of interface and slot/port information to view IS-IS information on that interface only.
- View information about IS-IS local update packets.

  EXEC Privilege mode

  ```
  debug isis local-updates [interface]
  ```

  To view specific information, enter the following optional parameter:

  - *`interface`*: Enter the type of interface and slot/port information to view IS-IS information on that interface only.
- View IS-IS SNP packets, include CSNPs and PSNPs.

  EXEC Privilege mode

  ```
  debug isis snp-packets [interface]
  ```

  To view specific information, enter the following optional parameter:

  - *`interface`*: Enter the type of interface and slot/port information to view IS-IS information on that interface only.
- View the events that triggered IS-IS shortest path first (SPF) events for debugging purposes.

  EXEC Privilege mode

  ```
  debug isis spf-triggers
  ```
- View sent and received LSPs.

  EXEC Privilege mode

  ```
  debug isis update-packets [interface]
  ```

  To view specific information, enter the following optional parameter:

  - *`interface`*: Enter the type of interface and slot/port information to view IS-IS information on that interface only.

Dell Networking OS displays debug messages on the console. To view which debugging commands are enabled, use the `show debugging` command in EXEC Privilege mode.

To disable a specific debug command, enter the keyword `no` then the `debug` command. For example, to disable debugging of IS-IS updates, use the `no debug isis updates-packets` command.

To disable all IS-IS debugging, use the `no debug isis` command.

To disable all debugging, use the `undebug all` command.

IS-IS Metric Styles

The following sections provide additional information about the IS-IS metric styles.

- Configuring the IS-IS Metric Style
- Configure Metric Values

Dell Networking OS supports the following IS-IS metric styles:

- narrow (supports only type, length, and value [TLV] up to 63)
- wide (supports TLV up to 16777215)
- transition (supports both narrow and wide and uses a TLV up to 63)
- narrow transition (accepts both narrow and wide and sends only narrow or old-style TLV)
- wide transition (accepts both narrow and wide and sends only wide or new-style TLV)

Configure Metric Values

For any level (Level-1, Level-2, or Level-1-2), the value range possible in the `isis metric` command in INTERFACE mode changes depending on the metric style.

The following describes the correct value range for the `isis metric` command.

Metric Style	Correct Value Range for the isis metric Command
wide	0 to 16777215
narrow	0 to 63

Metric Style	Correct Value Range for the isis metric Command
wide transition	0 to 16777215
narrow transition	0 to 63
transition	0 to 63

Maximum Values in the Routing Table

IS-IS metric styles support different cost ranges for the route. The cost range for the narrow metric style is 0 to 1023, while all other metric styles support a range of 0 to 0xFE000000.

Change the IS-IS Metric Style in One Level Only

By default, the IS-IS metric style is narrow. When you change from one IS-IS metric style to another, the IS-IS metric value (configured with the `isis metric` command) could be affected.

In the following scenarios, the IS-type is either Level-1 or Level-2 or Level-1-2 and the metric style changes.

Table 33. Metric Value When the Metric Style Changes

Beginning Metric Style	Final Metric Style	Resulting IS-IS Metric Value
wide	narrow	default value (10) if the original value is greater than 63. A message is sent to the console.
wide	transition	truncated value (the truncated value appears in the LSP only). The original `isis metric` value is displayed in the `show config` and `show running-config` commands and is used if you change back to transition metric style. NOTE: A truncated value is a value that is higher than 63, but set back to 63 because the higher value is not supported.
wide	narrow transition	default value (10) if the original value is greater than 63. A message is sent to the console.
wide	wide transition	original value
narrow	wide	original value
narrow	transition	original value
narrow	narrow transition	original value
narrow	wide transition	original value
transition	wide	original value

Beginning Metric Style	Final Metric Style	Resulting IS-IS Metric Value
transition	narrow	original value
transition	narrow	original value
transition	wide transition	original value
narrow transition	wide	original value
narrow transition	narrow	original value
narrow transition	wide transition	original value
narrow transition	transition	original value
wide transition	wide	original value
wide transition	narrow	default value (10) if the original value is greater than 63. A message is sent to the console.
wide transition	narrow transition	default value (10) if the original value is greater than 63. A message is sent to the console.
wide transition	transition	truncated value (the truncated value appears in the LSP only). The original `isis metric` value is displayed in the `show config` and `show running-config` commands and is used if you change back to transition metric style.

Moving to transition and then to another metric style produces different results.

Table 34. Metric Value when the Metric Style Changes Multiple Times

Beginning Metric Style	Next Metric Style	Resulting Metric Value	Next Metric Style	Final Metric Value
wide	transition	truncated value	wide	original value is recovered
wide transition	transition	truncated value	wide transition	original value is recovered
wide	transition	truncated value	narrow	default value (10). A message is sent to the logging buffer
wide transition	transition	truncated value	narrow transition	default value (10). A message is sent to the logging buffer

Leaks from One Level to Another

In the following scenarios, each IS-IS level is configured with a different metric style.

Table 35. Metric Value with Different Levels Configured with Different Metric Styles

Level-1 Metric Style	Level-2 Metric Style	Resulting Metric Value
narrow	wide	original value
narrow	wide transition	original value
narrow	narrow transition	original value
narrow	transition	original value
wide	narrow	truncated value
wide	narrow transition	truncated value
wide	wide transition	original value
wide	transition	truncated value
narrow transition	wide	original value
narrow transition	narrow	original value
narrow transition	wide transition	original value
narrow transition	transition	original value
transition	wide	original value
transition	narrow	original value
transition	wide transition	original value
transition	narrow transition	original value
wide transition	wide	original value
wide transition	narrow	truncated value
wide transition	narrow transition	truncated value
wide transition	transition	truncated value

Sample Configurations

The following configurations are examples for enabling IPv6 IS-IS. These examples are not comprehensive directions. They are intended to give you some guidance with typical configurations.

NOTE: Only one IS-IS process can run on the router, even if both IPv4 and IPv6 routing is being used.

You can copy and paste from these examples to your CLI. To support your own IP addresses, interfaces, names, and so on, be sure that you make the necessary changes.

NOTE: Whenever you make IS-IS configuration changes, clear the IS-IS process (re-started) using the `clear isis` command. The `clear isis` command must include the tag for the ISIS process. The following example shows the response from the router:

```
Dell#clear isis *
% ISIS not enabled.
Dell#clear isis 9999 *
```

You can configure IPv6 IS-IS routes in one of the following three different methods:

- **Congruent Topology** — You *must* configure both IPv4 and IPv6 addresses on the interface. Enable the `ip router isis` and `ipv6 router isis` commands on the interface. Enable the `wide-metrics` parameter in router isis configuration mode.
- **Multi-topology** — You *must* configure the IPv6 address. Configuring the IPv4 address is optional. You *must* enable the `ipv6 router isis` command on the interface. If you configure IPv4, also enable the `router isis` command. In router isis configuration mode, enable `multi-topology` under address-family ipv6 unicast.
- **Multi-topology Transition** — You *must* configure the IPv6 address. Configuring the IPv4 address is optional. You *must* enable the `ipv6 router isis` command on the interface. If you configure IPv4, also enable the `ip router isis` command. In router isis configuration mode, enable `multi-topology transition` under address-family ipv6 unicast.

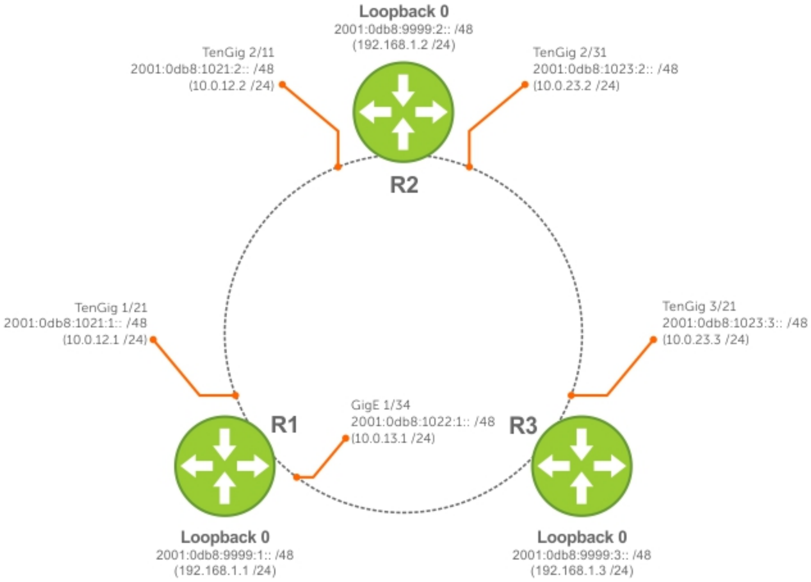

Loopback 0
2001:0db8:9999:2:: /48
(192.168.1.2 /24)
TenGig 2/11
2001:0db8:1021:2:: /48
(10.0.12.2 /24)
TenGig 2/31
2001:0db8:1023:2:: /48
(10.0.23.2 /24)
R2
TenGig 1/21
2001:0db8:1021:1:: /48
(10.0.12.1 /24)
TenGig 3/21
2001:0db8:1023:3:: /48
(10.0.23.3 /24)
R1
GigE 1/34
2001:0db8:1022:1:: /48
(10.0.13.1 /24)
R3
Loopback 0
2001:0db8:9999:1:: /48
(192.168.1.1 /24)
Loopback 0
2001:0db8:9999:3:: /48
(192.168.1.3 /24)

Figure 52. IPv6 IS-IS Sample Topography

IS-IS Sample Configuration — Congruent Topology

IS-IS Sample Configuration — Multi-topology

IS-IS Sample Configuration — Multi-topology Transition

The following is a sample configuration for enabling IPv6 IS-IS.

```
Dell(conf-if-te-3/17)#show config
!
interface TenGigabitEthernet 3/17
ip address 24.3.1.1/24
```

```
ipv6 address 24:3::1/76
ip router isis
ipv6 router isis
no shutdown
Dell (conf-if-te-3/17)#

Dell (conf-router_isis)#show config
!
router isis
metric-style wide level-1
metric-style wide level-2
net 34.0000.0000.AAAA.00
Dell (conf-router_isis)#


Dell (conf-if-te-3/17)#show config
!
interface TenGigabitEthernet 3/17
ipv6 address 24:3::1/76
ipv6 router isis
no shutdown
Dell (conf-if-te-3/17)#

Dell (conf-router_isis)#show config
!
router isis
net 34.0000.0000.AAAA.00
!
address-family ipv6 unicast
multi-topology
exit-address-family
Dell (conf-router_isis)#


Dell (conf-if-te-3/17)#show config
!
interface TenGigabitEthernet 3/17
ipv6 address 24:3::1/76
ipv6 router isis
no shutdown
Dell (conf-if-te-3/17)#

Dell (conf-router_isis)#show config
!
router isis
net 34.0000.0000.AAAA.00
!
address-family ipv6 unicast
multi-topology transition
exit-address-family
Dell (conf-router_isis)#
```

24

Link Aggregation Control Protocol (LACP)

Link aggregation control protocol (LACP) is supported on the S6000 platform.

Introduction to Dynamic LAGs and LACP

A link aggregation group (LAG), referred to as a *port channel* by Dell Networking OS, can provide both load-sharing and port redundancy across line cards. You can enable LAGs as static or dynamic.

The benefits and constraints are basically the same, as described in *Port Channel Interfaces* in the Interfaces chapter.

The unique benefit of a dynamic LAG is that its ports can toggle between participating in the LAG or acting as dedicated ports, whereas ports in a static LAG must be removed from the LAG in order to act alone.

The Dell Networking OS uses LACP to create dynamic LAGs. LACP provides a standardized means of exchanging information between two systems (also called Partner Systems) and automatically establishes the LAG between the systems. LACP permits the exchange of messages on a link to allow their LACP instances to:

- Reach an agreement on the identity of the LAG to which the link belongs.
- Move the link to that LAG.
- Enable the transmission and reception functions in an orderly manner.

The Dell Networking OS implementation of LACP is based on the standards specified in the IEEE 802.3: “Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications.”

LACP functions by constantly exchanging custom MAC protocol data units (PDUs) across local area network (LAN) Ethernet links. The protocol packets are only exchanged between ports that are configured as LACP capable.

Important Points to Remember

- LACP allows you to add members to a port channel (LAG) as long as it has no static members. Conversely, if the LAG already contains a statically defined member (the `channel-member` command), the `port-channel mode` command is not permitted.
- A static LAG cannot be created if a dynamic LAG using the selected number exists.
- No dual membership in static and dynamic LAGs:
  - If a physical interface is a part of a static LAG, the `port-channel-protocol lacp` command is rejected on that interface.
  - If a physical interface is a part of a dynamic LAG, it cannot be added as a member of a static LAG. The `channel-member gigabitethernet x/y` command is rejected in the static LAG interface for that physical interface.
- A dynamic LAG can be created with any type of configuration.

- There is a difference between the `shutdown` and `no interface port-channel` commands:
  - The `shutdown` command on LAG “xyz” disables the LAG and retains the user commands. However, the system does not allow the channel number “xyz” to be statically created.
  - The `no interface port-channel` *`channel-number`* command deletes the specified LAG, including a dynamically created LAG. This command removes all LACP-specific commands on the member interfaces. The interfaces are restored to a state that is ready to be configured.

    NOTE: There is no configuration on the interface because that condition is required for an interface to be part of a LAG.
- You can configure link dampening on individual members of a LAG.
- You can configure a maximum of up to 128 port-channels with eight members per channel.

LACP Modes

Dell Networking OS provides three modes for configuration of LACP — Off, Active, and Passive.

- **Off** — In this state, an interface is not capable of being part of a dynamic LAG. LACP does not run on any port that is configured to be in this state.
- **Active** — In this state, the interface is said to be in the “active negotiating state.” LACP runs on any link that is configured to be in this state. A port in Active state also automatically initiates negotiations with other ports by initiating LACP packets.
- **Passive** — In this state, the interface is not in an active negotiating state, but LACP runs on the link. A port in Passive state also responds to negotiation requests (from ports in Active state). Ports in Passive state respond to LACP packets.

Dell Networking OS supports LAGs in the following cases:

- A port in Active state can set up a port channel (LAG) with another port in Active state.
- A port in Active state can set up a LAG with another port in Passive state.

A port in Passive state cannot set up a LAG with another port in Passive state.

Configuring LACP Commands

If you configure aggregated ports with compatible LACP modes (Off, Active, Passive), LACP can automatically link them, as defined in IEEE 802.3, Section 43.
To configure LACP, use the following commands.

- Configure the system priority.

  CONFIGURATION mode

  ```
  [no] lacp system-priority priority-value
  ```

  The range is from 1 to 65535 (the higher the number, the lower the priority).

  The default is **32768**.
- Enable or disable LACP on any LAN port.

  INTERFACE mode

  ```
  [no] port-channel-protocol lacp
  ```

  The default is **LACP disabled**.

  This command creates context.

- Configure LACP mode.

  LACP mode

  ```
  [no] port-channel number mode [active | passive | off]
  ```

  - *`number`*: cannot statically contain any links.

  The default is **LACP active**.
- Configure port priority.

  LACP mode

  ```
  [no] lacp port-priority priority-value
  ```

  The range is from 1 to 65535 (the higher the number, the lower the priority).

  The default is **32768**.

LACP Configuration Tasks

The following are LACP configuration tasks.

- Creating a LAG
- Configuring the LAG Interfaces as Dynamic
- Setting the LACP Long Timeout
- Monitoring and Debugging LACP
- Configuring Shared LAG State Tracking

Creating a LAG

To create a dynamic port channel (LAG), use the following command. First you define the LAG and then the LAG interfaces.

- Create a dynamic port channel (LAG).

  CONFIGURATION mode

  ```
  interface port-channel
  ```
- Create a dynamic port channel (LAG).

  CONFIGURATION mode

  ```
  switchport
  ```

Example of Configuring a LAG Interface

```
Dell(conf)#interface port-channel 32
Dell(conf-if-po-32)#no shutdown
Dell(conf-if-po-32)#switchport
```

The LAG is in the default VLAN. To place the LAG into a non-default VLAN, use the `tagged` command on the LAG.

```
Dell(conf)#interface vlan 10
Dell(conf-if-vl-10)#tagged port-channel 32
```

Configuring the LAG Interfaces as Dynamic

After creating a LAG, configure the dynamic LAG interfaces.
To configure the dynamic LAG interfaces, use the following command.

- Configure the dynamic LAG interfaces.

  CONFIGURATION mode

  ```
  port-channel-protocol lacp
  ```

Example of the `port-channel-protocol lacp` Command

```
Dell(conf)#interface Gigabitethernet 3/15
Dell(conf-if-gi-3/15)#no shutdown
Dell(conf-if-gi-3/15)#port-channel-protocol lacp
Dell(conf-if-gi-3/15-lacp)#port-channel 32 mode active
...
Dell(conf)#interface Gigabitethernet 3/16
Dell(conf-if-gi-3/16)#no shutdown
Dell(conf-if-gi-3/16)#port-channel-protocol lacp
Dell(conf-if-gi-3/16-lacp)#port-channel 32 mode active
...
Dell(conf)#interface Gigabitethernet 4/15
Dell(conf-if-gi-4/15)#no shutdown
Dell(conf-if-gi-4/15)#port-channel-protocol lacp
Dell(conf-if-gi-4/15-lacp)#port-channel 32 mode active
...
Dell(conf)#interface Gigabitethernet 4/16
Dell(conf-if-gi-4/16)#no shutdown
Dell(conf-if-gi-4/16)#port-channel-protocol lacp
Dell(conf-if-gi-4/16-lacp)#port-channel 32 mode active
```

The port-channel 32 mode active command shown here may be successfully issued as long as there is no existing static channel-member configuration in LAG 32.

Setting the LACP Long Timeout

PDUs are exchanged between port channel (LAG) interfaces to maintain LACP sessions.
PDUs are transmitted at either a slow or fast transmission rate, depending upon the LACP timeout value. The timeout value is the amount of time that a LAG interface waits for a PDU from the remote system before bringing the LACP session down. The default timeout value is **1 second**. You can configure the default timeout value to be **30 seconds**. Invoking the longer timeout might prevent the LAG from flapping if the remote system is up but temporarily unable to transmit PDUs due to a system interruption.

NOTE: The 30-second timeout is available for dynamic LAG interfaces only. You can enter the `lacp long-timeout` command for static LAGs, but it has no effect.

To configure LACP long timeout, use the following command.

- Set the LACP timeout value to 30 seconds.

  CONFIG-INT-PO mode

  ```
  lacp long-timeout
  ```

Example of the `lacp long-timeout` and `show lacp` Commands

```
Dell(conf)# interface port-channel 32
Dell(conf-if-po-32)#no shutdown
```

```
Dell(conf-if-po-32)#switchport
Dell(conf-if-po-32)#lacp long-timeout
Dell(conf-if-po-32)#end
Dell# show lacp 32
Port-channel 32 admin up, oper up, mode lacp
Actor System ID: Priority 32768, Address 0001.e800.a12b
Partner System ID: Priority 32768, Address 0001.e801.45a5
Actor Admin Key 1, Oper Key 1, Partner Oper Key 1
LACP LAG 1 is an aggregatable link
A - Active LACP, B - Passive LACP, C - Short Timeout, D - Long Timeout
E - Aggregatable Link, F - Individual Link, G - IN_SYNC, H - OUT_OF_SYNC
I - Collection enabled, J - Collection disabled, K - Distribution enabled L -
Distribution disabled,
M - Partner Defaulted, N - Partner Non-defaulted, O - Receiver is in expired
state,
P - Receiver is not in expired state
Port Gi 10/6 is enabled, LACP is enabled and mode is lacp
Actor Admin: State ADEHJLMP Key 1 Priority 128
```

To view the PDU exchanges and the timeout value, use the `debug lacp` command. For more information, refer to Monitoring and Debugging LACP.

Monitoring and Debugging LACP

The system log (syslog) records faulty LACP actions.
To debug LACP, use the following command.

- Debug LACP, including configuration and events.

  EXEC mode

  ```
  [no] debug lacp [config | events | pdu [in | out | [interface [in | out]]]]
  ```

Shared LAG State Tracking

Shared LAG state tracking provides the flexibility to bring down a port channel (LAG) based on the operational state of another LAG.

At any time, only two LAGs can be a part of a group such that the fate (status) of one LAG depends on the other LAG.

As shown in the following illustration, the line-rate traffic from R1 destined for R4 follows the lowest-cost route via R2. Traffic is equally distributed between LAGs 1 and 2. If LAG 1 fails, all traffic from R1 to R4 flows across LAG 2 only. This condition over-subscribes the link and packets are dropped.

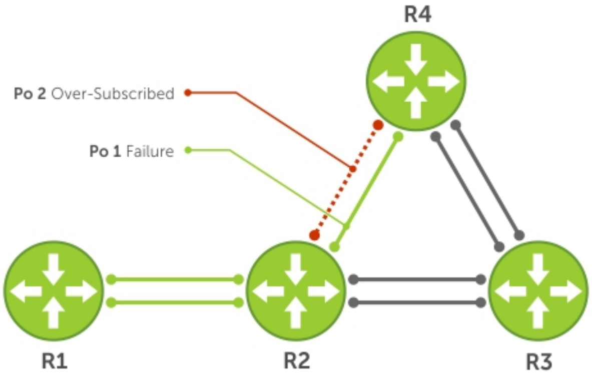


Figure 53. Shared LAG State Tracking

To avoid packet loss, redirect traffic through the next lowest-cost link (R3 to R4). Dell Networking OS has the ability to bring LAG 2 down if LAG 1 fails, so that traffic can be redirected. This redirection is what is meant by shared LAG state tracking. To achieve this functionality, you must group LAG 1 and LAG 2 into a single entity, called a failover group.

Configuring Shared LAG State Tracking

To configure shared LAG state tracking, you configure a failover group.

NOTE: If a LAG interface is part of a redundant pair, you cannot use it as a member of a failover group created for shared LAG state tracking.

1. Enter port-channel failover group mode.
   CONFIGURATION mode

   ```
   port-channel failover-group
   ```
2. Create a failover group and specify the two port-channels that will be members of the group.
   CONFIG-PO-FAILOVER-GRP mode

   ```
   group number port-channel number port-channel number
   ```

Example of LAGs in the Same Failover Group

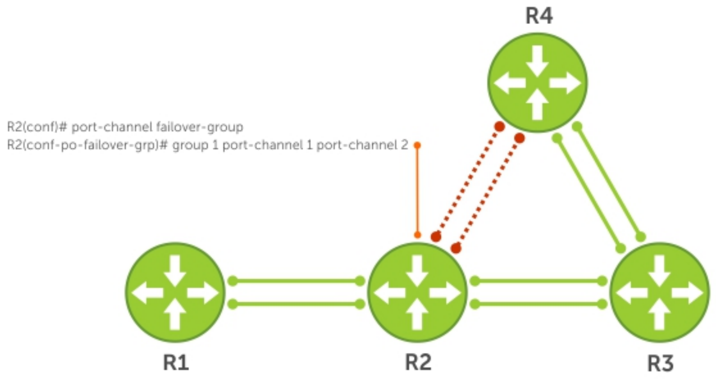
In the following example, LAGs 1 and 2 have been placed into to the same failover group.

```
Dell#config
Dell(conf)#port-channel failover-group
Dell(conf-po-failover-grp)#group 1 port-channel 1 port-channel 2
```

To view the failover group configuration, use the `show running-configuration po-failover-group` command.

```
Dell#show running-config po-failover-group
!
port-channel failover-group
group 1 port-channel 1 port-channel 2
```

As shown in the following illustration, LAGs 1 and 2 are members of a failover group. LAG 1 fails and LAG 2 is brought down after the failure. This effect is logged by Message 1, in which a console message declares both LAGs down at the same time.



```
R2(conf)# port-channel failover-group
R2(conf-po-failover-grp)# group 1 port-channel 1 port-channel 2
```

Figure 54. Configuring Shared LAG State Tracking

The following are shared LAG state tracking console messages:

- `2d1h45m: %RPM0-P:CP %IFMGR-5-OSTATE_DN: Changed interface state to down: Po 1`
- `2d1h45m: %RPM0-P:CP %IFMGR-5-OSTATE_DN: Changed interface state to down: Po 2`

To view the status of a failover group member, use the `show interface port-channel` command.

```
Dell#show interface port-channel 2
Port-channel 2 is up, line protocol is down (Failover-group 1 is down)
Hardware address is 00:01:e8:05:e8:4c, Current address is 00:01:e8:05:e8:4c
Interface index is 1107755010
Minimum number of links to bring Port-channel up is 1
Port-channel is part of failover-group 1
Internet address is not set
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 1000 Mbit
Members in this channel: Gi 1/17(U)
ARP type: ARPA, ARP Timeout 04:00:00
Last clearing of "show interface" counters 00:01:28
Queueing strategy: fifo
```

NOTE: The set of console messages shown above appear only if you configure shared LAG state tracking on that router (you can configure the feature on one or both sides of a link). For example, as previously shown, if you configured shared LAG state tracking on R2 only, no messages appear on R4 regarding the state of LAGs in a failover group.

Important Points about Shared LAG State Tracking

The following is more information about shared LAG state tracking.

- This feature is available for static and dynamic LAGs.
- Only a LAG can be a member of a failover group.
- You can configure shared LAG state tracking on one side of a link or on both sides.

- If a LAG that is part of a failover group is deleted, the failover group is deleted.
- If a LAG moves to the Down state due to this feature, its members may still be in the Up state.

LACP Basic Configuration Example

The screenshots in this section are based on the following example topology. Two routers are named ALPHA and BRAVO, and their hostname prompts reflect those names.

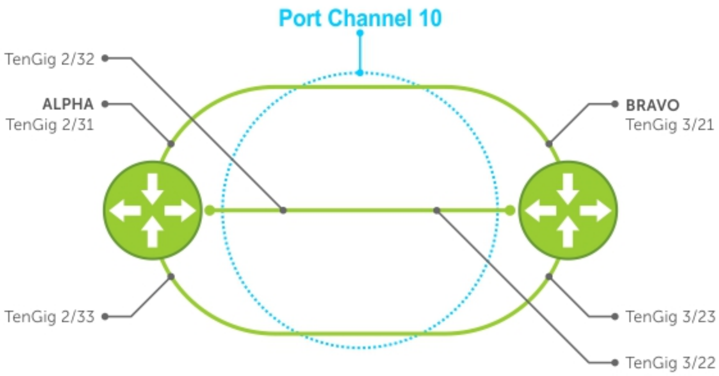


Figure 55. LACP Basic Configuration Example

Configure a LAG on ALPHA

The following example creates a LAG on ALPHA.

Example of Configuring a LAG

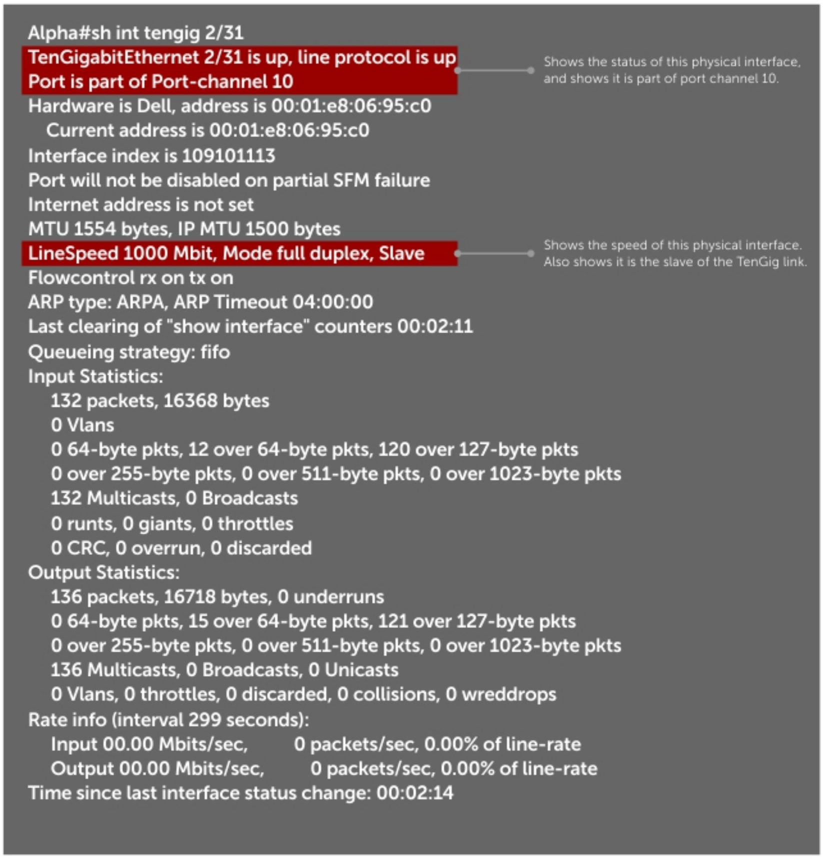
```
Alpha(conf)#interface port-channel 10
Alpha(conf-if-po-10)#no ip address
Alpha(conf-if-po-10)#switchport
Alpha(conf-if-po-10)#no shutdown
Alpha(conf-if-po-10)#show config
!
interface Port-channel 10
  no ip address
  switchport
  no shutdown
!
Alpha(conf-if-po-10)#
```

Example of Viewing a LAG Port Configuration

The following example inspects a LAG port configuration on ALPHA.

```
Alpha#sh int gig 2/31
GigabitEthernet 2/31 is up, line protocol is up
Port is part of Port-channel 10
Hardware is Force10Eth, address is 00:01:e8:06:95:c0
     Current address is 00:01:e8:06:95:c0
Interface Index is 109101113
Port will not be disabled on partial SFM failure
Internet address is not set
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 1000 Mbit, Mode full duplex, Slave
Flowcontrol rx on tx on
```

```
ARP type: ARPA, ARP Timeout 04:00:00
Last clearing of "show interface" counters 00:02:11
Queueing strategy: fifo
Input statistics:
     132 packets, 163668 bytes
     0 Vlans
     0 64-byte pkts, 12 over 64-byte pkts, 120 over 127-byte pkts
     0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
     132 Multicasts, 0 Broadcasts
     0 runts, 0 giants, 0 throttles
     0 CRC, 0 overrun, 0 discarded
Output Statistics
     136 packets, 16718 bytes, 0 underruns
     0 64-byte pkts, 15 over 64-byte pkts, 121 over 127-byte pkts
     0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
     136 Multicasts, 0 Broadcasts, 0 Unicasts
     0 Vlans, 0 throttles, 0 discarded, 0 collisions, 0 wreddrops
Rate info (interval 299 seconds):
     Input 00.00 Mbits/sec,0 packets/sec, 0.00% of line-rate
     Output 00.00 Mbits/sec,0 packets/sec, 0.00% of line-rate
     Time since last interface status change: 00:02:14
```

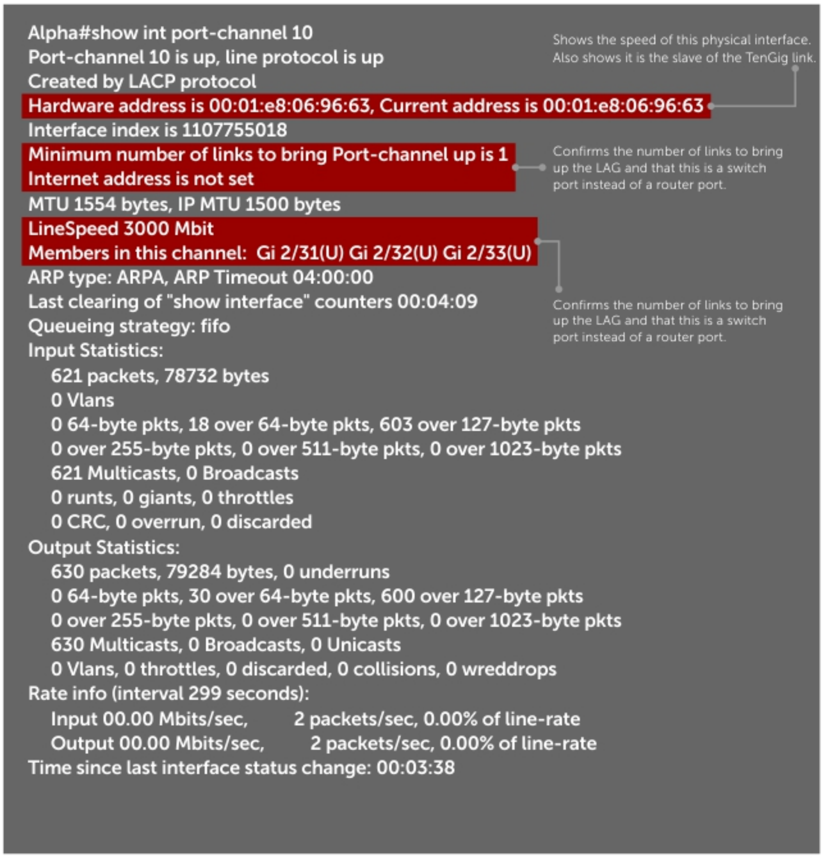


```
Alpha#sh int tengig 2/31
TenGigabitEthernet 2/31 is up, line protocol is up
Port is part of Port-channel 10
Hardware is Dell, address is 00:01:e8:06:95:c0
  Current address is 00:01:e8:06:95:c0
Interface index is 109101113
Port will not be disabled on partial SFM failure
Internet address is not set
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 1000 Mbit, Mode full duplex, Slave
Flowcontrol rx on tx on
ARP type: ARPA, ARP Timeout 04:00:00
Last clearing of "show interface" counters 00:02:11
Queueing strategy: fifo
Input Statistics:
   132 packets, 16368 bytes
   0 Vlans
   0 64-byte pkts, 12 over 64-byte pkts, 120 over 127-byte pkts
   0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
   132 Multicasts, 0 Broadcasts
   0 runts, 0 giants, 0 throttles
   0 CRC, 0 overrun, 0 discarded
Output Statistics:
   136 packets, 16718 bytes, 0 underruns
   0 64-byte pkts, 15 over 64-byte pkts, 121 over 127-byte pkts
   0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
   136 Multicasts, 0 Broadcasts, 0 Unicasts
   0 Vlans, 0 throttles, 0 discarded, 0 collisions, 0 wreddrops
Rate info (interval 299 seconds):
   Input 00.00 Mbits/sec,       0 packets/sec, 0.00% of line-rate
   Output 00.00 Mbits/sec,       0 packets/sec, 0.00% of line-rate
Time since last interface status change: 00:02:14
```

Shows the status of this physical interface, and shows it is part of port channel 10.

Shows the speed of this physical interface. Also shows it is the slave of the TenGig link.

Figure 56. Inspecting the LAG Configuration



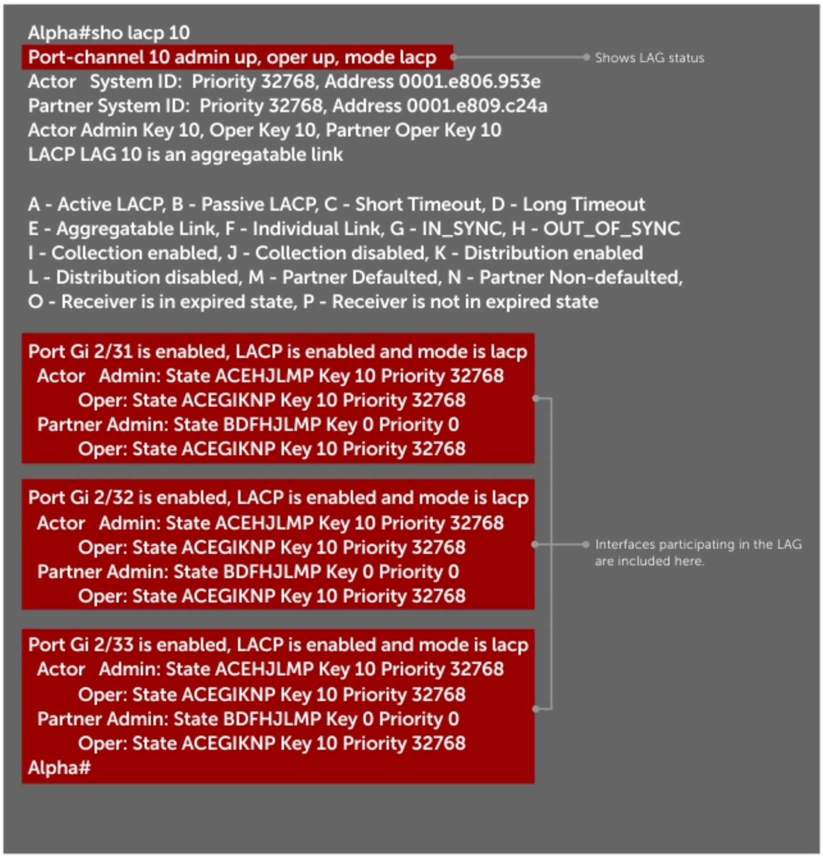
```
Alpha#show int port-channel 10
Port-channel 10 is up, line protocol is up
Created by LACP protocol
Hardware address is 00:01:e8:06:96:63, Current address is 00:01:e8:06:96:63
Interface index is 1107755018
Minimum number of links to bring Port-channel up is 1
Internet address is not set
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 3000 Mbit
Members in this channel:  Gi 2/31(U) Gi 2/32(U) Gi 2/33(U)
ARP type: ARPA, ARP Timeout 04:00:00
Last clearing of "show interface" counters 00:04:09
Queueing strategy: fifo
Input Statistics:
    621 packets, 78732 bytes
    0 Vlans
    0 64-byte pkts, 18 over 64-byte pkts, 603 over 127-byte pkts
    0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
    621 Multicasts, 0 Broadcasts
    0 runts, 0 giants, 0 throttles
    0 CRC, 0 overrun, 0 discarded
Output Statistics:
    630 packets, 79284 bytes, 0 underruns
    0 64-byte pkts, 30 over 64-byte pkts, 600 over 127-byte pkts
    0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
    630 Multicasts, 0 Broadcasts, 0 Unicasts
    0 Vlans, 0 throttles, 0 discarded, 0 collisions, 0 wreddrops
Rate info (interval 299 seconds):
    Input 00.00 Mbits/sec,        2 packets/sec, 0.00% of line-rate
    Output 00.00 Mbits/sec,        2 packets/sec, 0.00% of line-rate
Time since last interface status change: 00:03:38
```

Shows the speed of this physical interface. Also shows it is the slave of the TenGig link.

Confirms the number of links to bring up the LAG and that this is a switch port instead of a router port.

Confirms the number of links to bring up the LAG and that this is a switch port instead of a router port.

Figure 57. Inspecting Configuration of LAG 10 on ALPHA



```
Alpha#sho lacp 10
Port-channel 10 admin up, oper up, mode lacp
Actor   System ID:  Priority 32768, Address 0001.e806.953e
Partner System ID:  Priority 32768, Address 0001.e809.c24a
Actor Admin Key 10, Oper Key 10, Partner Oper Key 10
LACP LAG 10 is an aggregatable link

A - Active LACP, B - Passive LACP, C - Short Timeout, D - Long Timeout
E - Aggregatable Link, F - Individual Link, G - IN_SYNC, H - OUT_OF_SYNC
I - Collection enabled, J - Collection disabled, K - Distribution enabled
L - Distribution disabled, M - Partner Defaulted, N - Partner Non-defaulted,
O - Receiver is in expired state, P - Receiver is not in expired state

Port Gi 2/31 is enabled, LACP is enabled and mode is lacp
 Actor   Admin: State ACEHJLMP Key 10 Priority 32768
         Oper: State ACEGIKNP Key 10 Priority 32768
 Partner Admin: State BDFHJLMP Key 0 Priority 0
         Oper: State ACEGIKNP Key 10 Priority 32768

Port Gi 2/32 is enabled, LACP is enabled and mode is lacp
 Actor   Admin: State ACEHJLMP Key 10 Priority 32768
         Oper: State ACEGIKNP Key 10 Priority 32768
 Partner Admin: State BDFHJLMP Key 0 Priority 0
         Oper: State ACEGIKNP Key 10 Priority 32768

Port Gi 2/33 is enabled, LACP is enabled and mode is lacp
 Actor   Admin: State ACEHJLMP Key 10 Priority 32768
         Oper: State ACEGIKNP Key 10 Priority 32768
 Partner Admin: State BDFHJLMP Key 0 Priority 0
         Oper: State ACEGIKNP Key 10 Priority 32768
Alpha#
```

Figure 58. Verifying LAG 10 Status on ALPHA Using the show lacp Command

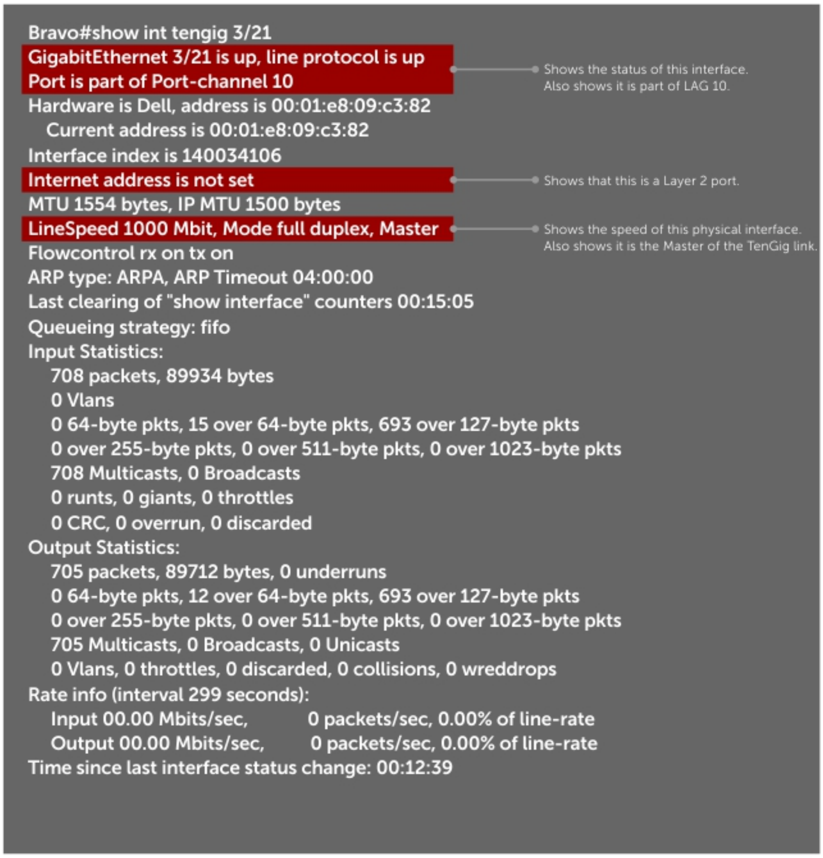
Summary of the LAG Configuration on Alpha

```
Alpha(conf-if-po-10)#int gig 2/31
Alpha(conf-if-gi-2/31)#no ip address
Alpha(conf-if-gi-2/31)#no switchport
Alpha(conf-if-gi-2/31)#shutdown
Alpha(conf-if-gi-2/31)#port-channel-protocol lacp
Alpha(conf-if-gi-2/31-lacp)#port-channel 10 mode active
Alpha(conf-if-gi-2/31-lacp)#no shut
Alpha(conf-if-gi-2/31)#show config

!
interface GigabitEthernet 2/31
  no ip address
!
  port-channel-protocol LACP
   port-channel 10 mode active
  no shutdown
!
Alpha(conf-if-gi-2/31)#

interface Port-channel 10
no ip address
switchport
no shutdown
```

```
interface GigabitEthernet 2/31
no ip address
```

Summary of the LAG Configuration on Bravo

```
Bravo(conf-if-gi-3/21)#int port-channel 10
Bravo(conf-if-po-10)#no ip add
Bravo(conf-if-po-10)#switch
Bravo(conf-if-po-10)#no shut
Bravo(conf-if-po-10)#show config
!
interface Port-channel 10
  no ip address
  switchport
  no shutdown
!
Bravo(conf-if-po-10)#exit

Bravo(conf)#int gig 3/21
Bravo(conf)#no ip address
Bravo(conf)#no switchport
Bravo(conf)#shutdown
Bravo(conf-if-gi-3/21)#port-channel-protocol lacp
Bravo(conf-if-gi-3/21-lacp)#port-channel 10 mode active
Bravo(conf-if-gi-3/21-lacp)#no shut
Bravo(conf-if-gi-3/21)#end

!
interface GigabitEthernet 3/21
  no ip address
!
  port-channel-protocol LACP
   port-channel 10 mode active
  no shutdown
Bravo(conf-if-gi-3/21)#end

int port-channel 10
no ip address
switchport
no shutdown
show config

int gig 3/21
no ip address
```



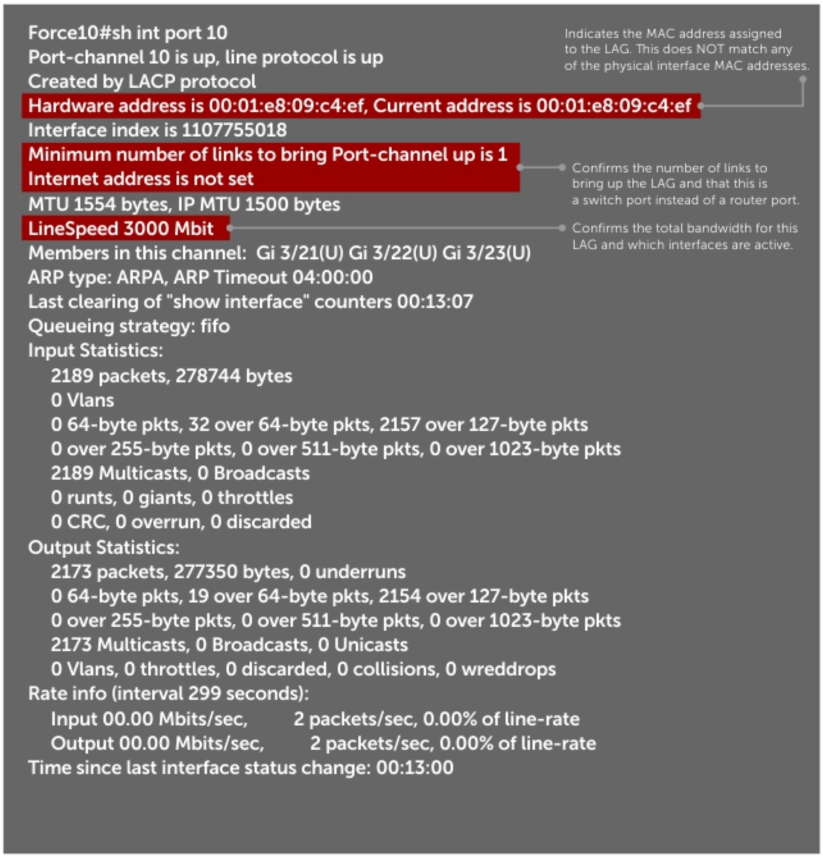
```
Bravo#show int tengig 3/21
GigabitEthernet 3/21 is up, line protocol is up
Port is part of Port-channel 10
Hardware is Dell, address is 00:01:e8:09:c3:82
  Current address is 00:01:e8:09:c3:82
Interface index is 140034106
Internet address is not set
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 1000 Mbit, Mode full duplex, Master
Flowcontrol rx on tx on
ARP type: ARPA, ARP Timeout 04:00:00
Last clearing of "show interface" counters 00:15:05
Queueing strategy: fifo
Input Statistics:
   708 packets, 89934 bytes
   0 Vlans
   0 64-byte pkts, 15 over 64-byte pkts, 693 over 127-byte pkts
   0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
   708 Multicasts, 0 Broadcasts
   0 runts, 0 giants, 0 throttles
   0 CRC, 0 overrun, 0 discarded
Output Statistics:
   705 packets, 89712 bytes, 0 underruns
   0 64-byte pkts, 12 over 64-byte pkts, 693 over 127-byte pkts
   0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
   705 Multicasts, 0 Broadcasts, 0 Unicasts
   0 Vlans, 0 throttles, 0 discarded, 0 collisions, 0 wreddrops
Rate info (interval 299 seconds):
   Input 00.00 Mbits/sec,          0 packets/sec, 0.00% of line-rate
   Output 00.00 Mbits/sec,         0 packets/sec, 0.00% of line-rate
Time since last interface status change: 00:12:39
```

Shows the status of this interface.
Also shows it is part of LAG 10.

Shows that this is a Layer 2 port.

Shows the speed of this physical interface.
Also shows it is the Master of the TenGig link.

Figure 59. Inspecting a LAG Port on BRAVO Using the show interface Command



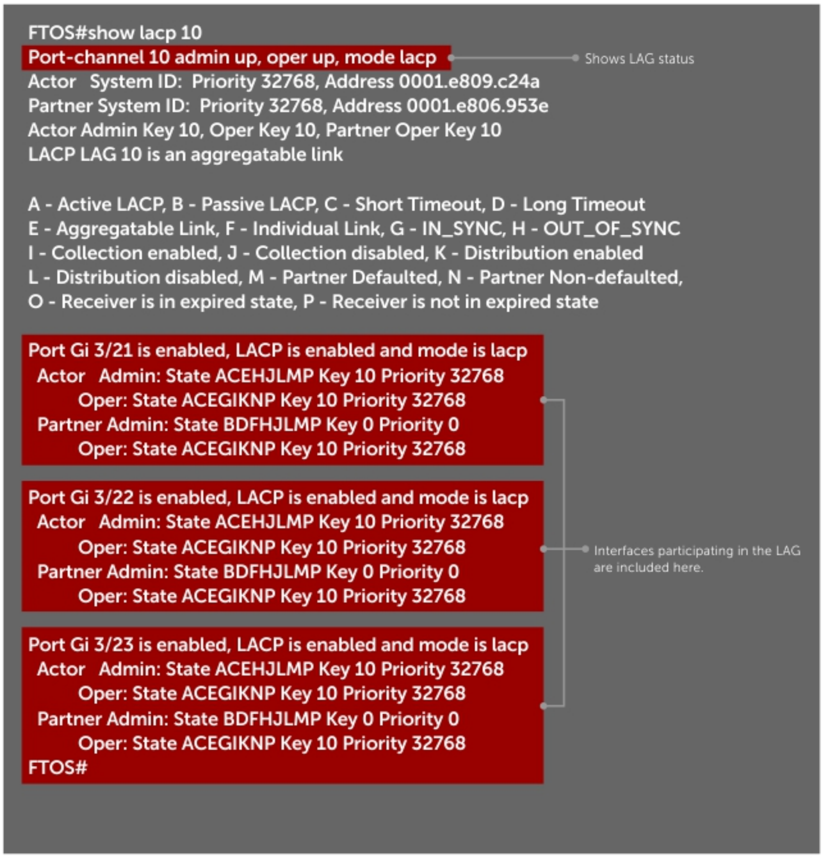
```
Force10#sh int port 10
Port-channel 10 is up, line protocol is up
Created by LACP protocol
Hardware address is 00:01:e8:09:c4:ef, Current address is 00:01:e8:09:c4:ef
Interface index is 1107755018
Minimum number of links to bring Port-channel up is 1
Internet address is not set
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed 3000 Mbit
Members in this channel:  Gi 3/21(U) Gi 3/22(U) Gi 3/23(U)
ARP type: ARPA, ARP Timeout 04:00:00
Last clearing of "show interface" counters 00:13:07
Queueing strategy: fifo
Input Statistics:
    2189 packets, 278744 bytes
    0 Vlans
    0 64-byte pkts, 32 over 64-byte pkts, 2157 over 127-byte pkts
    0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
    2189 Multicasts, 0 Broadcasts
    0 runts, 0 giants, 0 throttles
    0 CRC, 0 overrun, 0 discarded
Output Statistics:
    2173 packets, 277350 bytes, 0 underruns
    0 64-byte pkts, 19 over 64-byte pkts, 2154 over 127-byte pkts
    0 over 255-byte pkts, 0 over 511-byte pkts, 0 over 1023-byte pkts
    2173 Multicasts, 0 Broadcasts, 0 Unicasts
    0 Vlans, 0 throttles, 0 discarded, 0 collisions, 0 wreddrops
Rate info (interval 299 seconds):
    Input 00.00 Mbits/sec,        2 packets/sec, 0.00% of line-rate
    Output 00.00 Mbits/sec,        2 packets/sec, 0.00% of line-rate
Time since last interface status change: 00:13:00
```

Indicates the MAC address assigned to the LAG. This does NOT match any of the physical interface MAC addresses.

Confirms the number of links to bring up the LAG and that this is a switch port instead of a router port.

Confirms the total bandwidth for this LAG and which interfaces are active.

Figure 60. Inspecting LAG 10 Using the show interfaces port-channel Command



```
FTOS#show lacp 10
Port-channel 10 admin up, oper up, mode lacp
Actor   System ID:  Priority 32768, Address 0001.e809.c24a
Partner System ID:  Priority 32768, Address 0001.e806.953e
Actor Admin Key 10, Oper Key 10, Partner Oper Key 10
LACP LAG 10 is an aggregatable link

A - Active LACP, B - Passive LACP, C - Short Timeout, D - Long Timeout
E - Aggregatable Link, F - Individual Link, G - IN_SYNC, H - OUT_OF_SYNC
I - Collection enabled, J - Collection disabled, K - Distribution enabled
L - Distribution disabled, M - Partner Defaulted, N - Partner Non-defaulted,
O - Receiver is in expired state, P - Receiver is not in expired state

Port Gi 3/21 is enabled, LACP is enabled and mode is lacp
 Actor   Admin: State ACEHJLMP Key 10 Priority 32768
         Oper: State ACEGIKNP Key 10 Priority 32768
 Partner Admin: State BDFHJLMP Key 0 Priority 0
         Oper: State ACEGIKNP Key 10 Priority 32768

Port Gi 3/22 is enabled, LACP is enabled and mode is lacp
 Actor   Admin: State ACEHJLMP Key 10 Priority 32768
         Oper: State ACEGIKNP Key 10 Priority 32768
 Partner Admin: State BDFHJLMP Key 0 Priority 0
         Oper: State ACEGIKNP Key 10 Priority 32768

Port Gi 3/23 is enabled, LACP is enabled and mode is lacp
 Actor   Admin: State ACEHJLMP Key 10 Priority 32768
         Oper: State ACEGIKNP Key 10 Priority 32768
 Partner Admin: State BDFHJLMP Key 0 Priority 0
         Oper: State ACEGIKNP Key 10 Priority 32768
FTOS#
```

Shows LAG status

Interfaces participating in the LAG are included here.

Figure 61. Inspecting the LAG Status Using the show lacp command

The point-to-point protocol (PPP) is a connection-oriented protocol that enables layer two links over various different physical layer connections. It is supported on both synchronous and asynchronous lines, and can operate in Half-Duplex or Full-Duplex mode. It was designed to carry IP traffic but is general enough to allow any type of network layer datagram to be sent over a PPP connection. As its name implies, it is for point-to-point connections between exactly two devices, and assumes that frames are sent and received in the same order.

25

Layer 2

Layer 2 features are supported on the S6000 platform.

Manage the MAC Address Table

Dell Networking OS provides the following management activities for the MAC address table.

- Clearing the MAC Address Table
- Setting the Aging Time for Dynamic Entries
- Configuring a Static MAC Address
- Displaying the MAC Address Table

Clearing the MAC Address Table

You may clear the MAC address table of dynamic entries.
To clear a MAC address table, use the following command.

- Clear a MAC address table of dynamic entries.

  EXEC Privilege mode

  ```
  clear mac-address-table {dynamic | sticky} {address | all | interface | vlan}
  ```

  - *`address`*: deletes the specified entry.
  - `all`: deletes all dynamic entries.
  - `interface`: deletes all entries for the specified interface.
  - `vlan`: deletes all entries for the specified VLAN.

Setting the Aging Time for Dynamic Entries

Learned MAC addresses are entered in the table as dynamic entries, which means that they are subject to aging.
For any dynamic entry, if no packet arrives on the switch with the MAC address as the source or destination address within the timer period, the address is removed from the table. The default aging time is **1800 seconds**.

To disable a MAC address and specify an aging time, use the following commands.

- Disable MAC address aging for all dynamic entries.

  CONFIGURATION mode

  ```
  mac-address-table aging-time 0
  ```
- Specify an aging time.

  CONFIGURATION mode

  ```
  mac-address-table aging-time seconds
  ```

The range is from 10 to 1000000.

Configuring a Static MAC Address

A static entry is one that is not subject to aging. Enter static entries manually.
To create a static MAC address entry, use the following command.

- Create a static MAC address entry in the MAC address table.

  CONFIGURATION mode

  ```
  mac-address-table static
  ```

Displaying the MAC Address Table

To display the MAC address table, use the following command.

- Display the contents of the MAC address table.

  EXEC Privilege mode

  ```
  show mac-address-table [address | aging-time [vlan vlan-id]| count | dynamic
  | interface | static | vlan]
  ```

  - `address`: displays the specified entry.
  - `aging-time`: displays the configured aging-time.
  - `count`: displays the number of dynamic and static entries for all VLANs, and the total number of entries.
  - `dynamic`: displays only dynamic entries.
  - `interface`: displays only entries for the specified interface.
  - `static`: displays only static entries.
  - `vlan`: displays only entries for the specified VLAN.

MAC Learning Limit

MAC address learning limit is a method of port security on Layer 2 port-channel and physical interfaces, and VLANs. It allows you to set an upper limit on the number of MAC addresses that learned on an interface/VLAN. After the limit is reached, the system drops all traffic from a device with an unlearned MAC address.

This section describes the following:

- Setting the MAC Learning Limit
- mac learning-limit Dynamic
- mac learning-limit mac-address-sticky
- mac learning-limit station-move
- Learning Limit Violation Actions
- Setting Station Move Violation Actions
- Recovering from Learning Limit and Station Move Violations

Dell Networking OS Behavior: When configuring the MAC learning limit on a port or VLAN, the configuration is accepted (becomes part of `running-config` and `show mac learning-limit` `interface`) before the system verifies that sufficient CAM space exists. If the CAM check fails, a message is displayed:

```
%E90MH:5 %ACL_AGENT-2-ACL_AGENT_LIST_ERROR: Unable to apply access-list Mac-
Limit on GigabitEthernet 5/84
```

In this case, the configuration is still present in the `running-config` and `show` output. Remove the configuration before re-applying a MAC learning limit with a lower value. Also, ensure that you can view the Syslog messages on your session.

NOTE: The CAM-check failure message beginning in Dell Networking OS version 8.3.1.0 is different from versions 8.2.1.1 and earlier, which read:

```
% Error: ACL returned error
% Error: Remove existing limit configuration if it was configured before
```

Setting the MAC Learning Limit

To set a MAC learning limit on an interface, use the following command.

- Specify the number of MAC addresses that the system can learn off a Layer 2 interface.
  INTERFACE mode

  ```
  mac learning-limit address_limit
  ```

  Three options are available with the `mac learning-limit` command:

  - `dynamic`
  - `no-station-move`
  - `station-move`

  NOTE: An SNMP trap is available for `mac learning-limit station-move`. No other SNMP traps are available for MAC Learning Limit, including limit violations.

mac learning-limit Dynamic

The MAC address table is stored on the Layer 2 forwarding information base (FIB) region of the CAM.

The Layer 2 FIB region allocates space for static MAC address entries and dynamic MAC address entries. When you enable MAC learning limit, entries created on this port are static by default. When you configure the `dynamic` option, learned MAC addresses are stored in the dynamic region and are subject to aging. Entries created before this option is set are not affected.

Dell Networking OS Behavior: If you do not configure the `dynamic` option, the systems do not detect station moves in which a MAC address learned off of a MAC-limited port is learned on another port on same line card. Therefore, Dell Networking OS does not take any configured station-move violation action. When a MAC address is relearned on any other line card (any line card except the one to which the original MAC-limited port belongs), the station-move is detected and the system takes the configured the violation action.

mac learning-limit mac-address-sticky

Using sticky MAC addresses allows you to associate a specific port with MAC addresses from trusted devices. If you enable sticky MAC, the specified port retains any dynamically-learned addresses and prevents them from being transferred or learned on other ports.

If you configure `mac-learning-limit` and you enabled sticky MAC, all dynamically-learned addresses are converted to sticky MAC addresses for the selected port. Any new MAC addresses learned on this port is converted to sticky MAC addresses.

To save all sticky MAC addresses into a configuration file that can be used as a startup configuration file, use the `write config` command. If the number of existing MAC addresses is fewer than the configured mac learn limit, any additional MAC addresses are converted to sticky MACs on that interface. To remove all sticky MAC addresses from the running config file, disable sticky MAC and use the `write config` command.

When you enable sticky mac on an interface, dynamically-learned MAC addresses do not age, even if you enabled `mac-learning-limit dynamic`. If you configured `mac-learning-limit` and `mac-learning-limit dynamic` and you disabled sticky MAC, any dynamically-learned MAC addresses ages.

mac learning-limit station-move

The `mac learning-limit station-move` command is available on the S6000 platform.

The `station-move` option, allows a MAC address already in the table to be learned off of another interface. For example, if you disconnect a network device from one interface and reconnect it to another interface, the MAC address is learned on the new interface. When the system detects this “station move,” the system clears the entry learned on the original interface and installs a new entry on the new interface.

mac learning-limit no-station-move

The `no-station-move` option, also known as “sticky MAC,” provides additional port security by preventing a station move.
When you configure this option, the first entry in the table is maintained instead of creating an entry on the new interface. `no-station-move` is the default behavior. Entries created before you set this option are not affected.

To display a list of all interfaces with a MAC learning limit, use the following command.

Display a list of all interfaces with a MAC learning limit.
EXEC Privilege mode

```
show mac learning-limit
```

Dell Networking OS Behavior: The systems do not generate a station-move violation log entry for physical interfaces or port-channels when you configure `mac learning-limit` or when you configure `mac learning-limit station-move-violation log`. Dell Networking OS detects a station-move violation only when you configure `mac learning-limit dynamic` and logs the violation only when you configure the `mac learning-limit station-move-violation log`, as shown in the following example.

```
Dell(conf-if-gi-1/1)#show config
!
interface GigabitEthernet 1/1
```

```
no ip address
switchport
mac learning-limit 1 dynamic no-station-move
mac learning-limit station-move-violation log
no shutdown
```

Learning Limit Violation Actions

Learning limit violation actions are supported only on the S6000 platform.
To configure the system to take an action when the MAC learning limit is reached on an interface and a new address is received using one the following options with the `mac learning-limit` command, use the following commands.

- Generate a system log message when the MAC learning limit is exceeded.

  INTERFACE mode

  ```
  learn-limit-violation log
  ```
- Shut down the interface and generate a system log message when the MAC learning limit is exceeded.

  INTERFACE mode

  ```
  learn-limit-violation shutdown
  ```

Setting Station Move Violation Actions

Station move violation actions are supported only on the S6000 platform.
`no-station-move` is the default behavior. You can configure the system to take an action if a station move occurs using one the following options with the `mac learning-limit` command.

To display a list of interfaces configured with MAC learning limit or station move violation actions, use the following commands.

- Generate a system log message indicating a station move.

  INTERFACE mode

  ```
  station-move-violation log
  ```
- Shut down the first port to learn the MAC address.

  INTERFACE mode

  ```
  station-move-violation shutdown-original
  ```
- Shut down the second port to learn the MAC address.

  INTERFACE mode

  ```
  station-move-violation shutdown-offending
  ```
- Shut down both the first and second port to learn the MAC address.

  INTERFACE mode

  ```
  station-move-violation shutdown-both
  ```
- Display a list of all of the interfaces configured with MAC learning limit or station move violation.

  CONFIGURATION mode

  ```
  show mac learning-limit violate-action
  ```

Recovering from Learning Limit and Station Move Violations

After a learning-limit or station-move violation shuts down an interface, you must manually reset it.
To reset the learning limit, use the following commands.

NOTE: Alternatively, you can reset the interface by shutting it down using the `shutdown` command and then re-enabling it using the `no shutdown` command.

- Reset interfaces in the ERR_Disabled state caused by a learning limit violation or station move violation.

  EXEC Privilege mode

  ```
  mac learning-limit reset
  ```
- Reset interfaces in the ERR_Disabled state caused by a learning limit violation.

  EXEC Privilege mode

  ```
  mac learning-limit reset learn-limit-violation [interface | all]
  ```
- Reset interfaces in the ERR_Disabled state caused by a station move violation.

  EXEC Privilege mode

  ```
  mac learning-limit reset station-move-violation [interface | all]
  ```

NIC Teaming

Network interface controller (NIC) teaming is available on the Z-Series platform.

NIC teaming is a feature that allows multiple network interface cards in a server to be represented by one MAC address and one IP address in order to provide transparent redundancy, balancing, and to fully utilize network adapter resources.

The following illustration shows a topology where two NICs have been teamed together. In this case, if the primary NIC fails, traffic switches to the secondary NIC because they are represented by the same set of addresses.

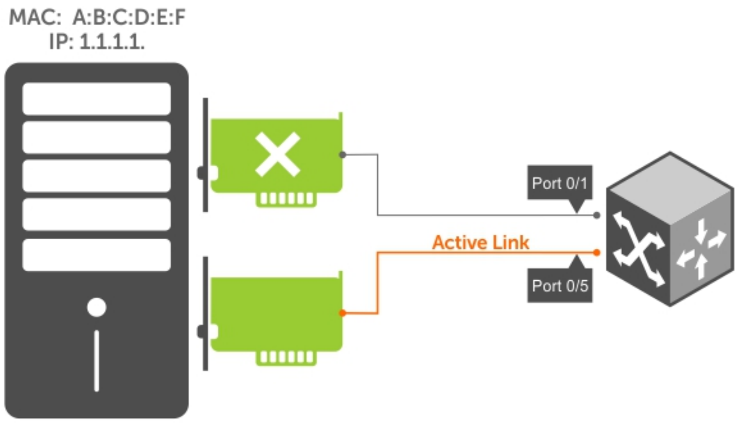


Figure 62. Redundant NICs with NIC Teaming

When you use NIC teaming, consider that the server MAC address is originally learned on Port 0/1 of the switch (shown in the following) and Port 0/5 is the failover port. When the NIC fails, the system automatically sends an ARP request for the gateway or host NIC to resolve the ARP and refresh the egress interface. When the ARP is resolved, the same MAC address is learned on the same port where the ARP is resolved (in the previous example, this location is Port 0/5 of the switch). To ensure that the MAC address is disassociated with one port and re-associated with another port in the ARP table, configure the `mac-address-table station-move refresh-arp` command on the Dell Networking switch at the time that NIC teaming is being configured on the server.

NOTE: If you do not configure the `mac-address-table station-move refresh-arp` command, traffic continues to be forwarded to the failed NIC until the ARP entry on the switch times out.

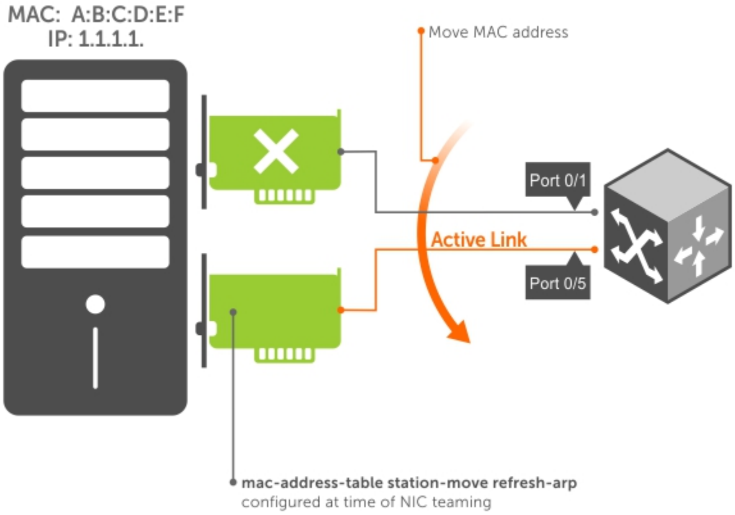

MAC: A:B:C:D:E:F
IP: 1.1.1.1.
Move MAC address
Port 0/1
Active Link
Port 0/5
mac-address-table station-move refresh-arp
configured at time of NIC teaming

Figure 63. Configuring the mac-address-table station-move refresh-arp Command

Configure Redundant Pairs

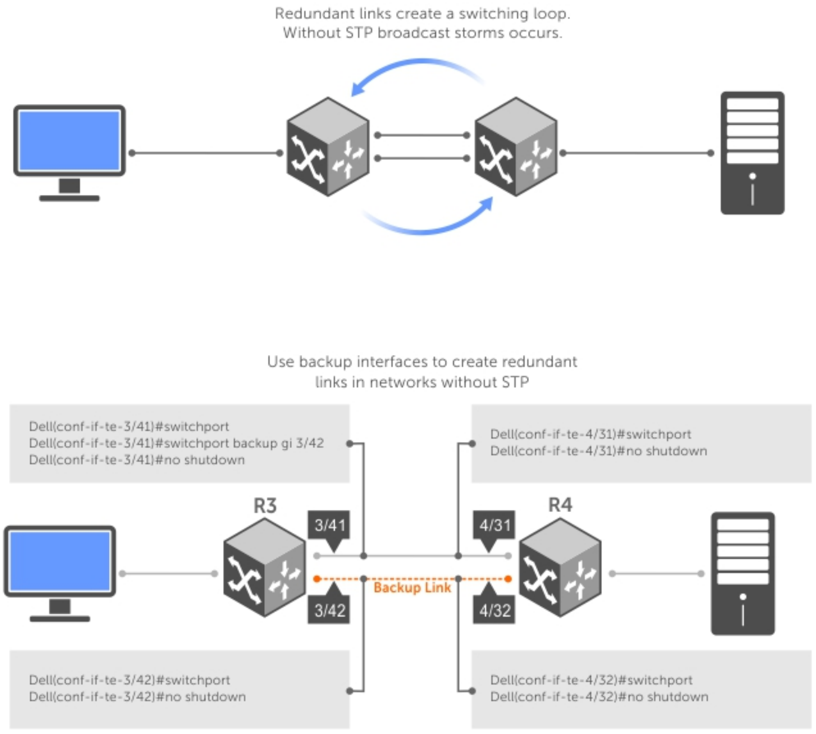
Configuring redundant pairs is supported on the S6000 platform.

Networks that employ switches that do not support the spanning tree protocol (STP) — for example, networks with digital subscriber line access multiplexers (DSLAM) — cannot have redundant links between switches because they create switching loops (as shown in the following illustration). The redundant pairs feature allows you to create redundant links in networks that do not use STP by configuring backup interfaces for the interfaces on either side of the primary link.

NOTE: For more information about STP, refer to Spanning Tree Protocol (STP).

Assign a backup interface to an interface using the `switchport backup` command. The backup interface remains in a Down state until the primary fails, at which point it transitions to Up state. If the primary interface fails, and later comes up, it becomes the backup interface for the redundant pair. Dell Networking OS supports Gigabit, 10 Gigabit, and 40-Gigabit interfaces as backup interfaces.

Apply all other configurations to each interface in the redundant pair such that their configurations are *identical*, so that transition to the backup interface in the event of a failure is transparent to rest of the network.



Redundant links create a switching loop.
Without STP broadcast storms occurs.

Use backup interfaces to create redundant
links in networks without STP

```
Dell(conf-if-te-3/41)#switchport
Dell(conf-if-te-3/41)#switchport backup gi 3/42
Dell(conf-if-te-3/41)#no shutdown
```

```
Dell(conf-if-te-4/31)#switchport
Dell(conf-if-te-4/31)#no shutdown
```

```
Dell(conf-if-te-3/42)#switchport
Dell(conf-if-te-3/42)#no shutdown
```

```
Dell(conf-if-te-4/32)#switchport
Dell(conf-if-te-4/32)#no shutdown
```

Figure 64. Configuring Redundant Layer 2 Pairs without Spanning Tree

You configure a redundant pair by assigning a backup interface to a primary interface with the `switchport backup interface` command. Initially, the primary interface is active and transmits traffic and the backup interface remains down. If the primary fails for any reason, the backup transitions to an active Up state. If the primary interface fails and later comes back up, it remains as the backup interface for the redundant pair.

Dell Networking OS supports only Gigabit, 10 Gigabit, and 40-Gigabit ports and port channels as primary/backup interfaces in redundant pairs. (A port channel is also referred to as a link aggregation group (LAG). For more information, refer to Interfaces) If the interface is a member link of a LAG, the following primary/backup interfaces are also supported:

- primary interface is a physical interface, the backup interface can be a physical interface
- primary interface is a physical interface, the backup interface can be a static or dynamic LAG
- primary interface is a static or dynamic LAG, the backup interface can be a physical interface
- primary interface is a static or dynamic LAG, the backup interface can be a static or dynamic LAG

In a redundant pair, any combination of physical and port-channel interfaces is supported as the two interfaces in a redundant pair. For example, you can configure a static (without LACP) or dynamic (with

LACP) port-channel interface as either the primary or backup link in a redundant pair with a physical interface.

To ensure that existing network applications see no difference when a primary interface in a redundant pair transitions to the backup interface, be sure to apply identical configurations of other traffic parameters to each interface.

If you remove an interface in a redundant link (remove the line card of a physical interface or delete a port channel with the `no interface port-channel` command), the redundant pair configuration is also removed.

Important Points about Configuring Redundant Pairs

- You may not configure any interface to be a backup for more than one interface, no interface can have more than one backup, and a backup interface may not have a backup interface.
- The active or backup interface may not be a member of a LAG.
- The active and standby do not have to be of the same type (1G, 10G, and so on).
- You may not enable any Layer 2 protocol on any interface of a redundant pair or to ports connected to them.

As shown in the previous illustration, interface 3/41 is a backup interface for 3/42, and 3/42 is in the Down state. If 3/41 fails, 3/42 transitions to the Up state, which makes the backup link active. A message similar to the following message appears whenever you configure a backup port.

```
02:28:04: %RPM0-P:CP %IFMGR-5-L2BKUP_WARN: Do not run any Layer2 protocols on
Gi 3/41
and Gi 3/42
02:28:04: %RPM0-P:CP %IFMGR-5-OSTATE_DN: Changed interface state to down: Gi 3/42
02:28:04: %RPM0-P:CP %IFMGR-5-STATE_ACT_STBY: Changed interface state to
standby: Gi
3/42
```

Example of Configuring Redundant Layer 2 Pairs

```
Dell(conf-if-range-gi-3/41-42)#switchport backup interface GigabitEthernet 3/42
Dell(conf-if-range-gi-3/41-42)#show config
!
interface GigabitEthernet 3/41
  no ip address
  switchport
  switchport backup interface GigabitEthernet 3/42
  no shutdown
!
interface GigabitEthernet 3/42
  no ip address
  switchport
  no shutdown
Dell(conf-if-range-gi-3/41-42)#
Dell(conf-if-range-gi-3/41-42)#do show ip int brief | find 3/41
GigabitEthernet 3/41   unassigned         YES Manual up       up
GigabitEthernet 3/42   unassigned         NO Manual up        down
[output omitted]
Dell(conf-if-range-gi-3/41-42)#interface gig 3/41
Dell(conf-if-gi-3/41)#shutdown
00:24:53: %RPM0-P:CP %IFMGR-5-ASTATE_DN: Changed interface Admin state to down:
Gi 3/41
Dell(conf-if-gi-3/41)#00:24:55: %RPM0-P:CP %IFMGR-5-OSTATE_DN: Changed
interface state to
down: Gi 3/41
00:24:55: %RPM0-P:CP %IFMGR-5-INACTIVE: Changed Vlan interface state to
```

```
inactive: Vl 1
00:24:55: %RPM0-P:CP %IFMGR-5-OSTATE_UP: Changed interface state to up: Gi 3/42
00:24:55: %RPM0-P:CP %IFMGR-5-ACTIVE: Changed Vlan interface state to active:
Vl 1
00:24:55: %RPM0-P:CP %IFMGR-5-STATE_STBY_ACT: Changed interface state from standby to active:
Gi 3/42

Dell(conf-if-gi-3/41)#do show ip int brief | find 3/41
GigabitEthernet 3/41   unassigned      NO Manual administratively down down
GigabitEthernet 3/42   unassigned      YES Manual up                   up
[output omitted]
```

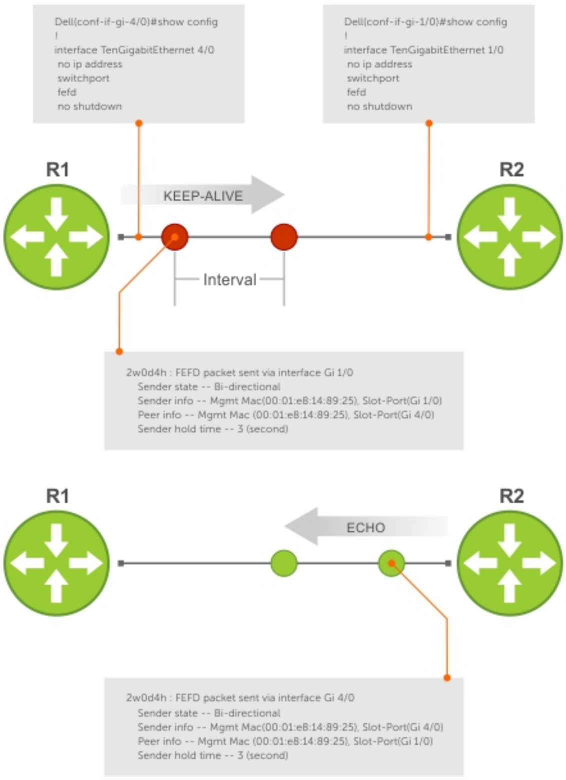
Example of Configuring Redundant Pairs on a Port-Channel ()

```
Dell#show interfaces port-channel brief
Codes: L - LACP Port-channel

  LAG  Mode  Status  Uptime     Ports
  1    L2    up      00:08:33   Te 0/0 (Up)
  2    L2    up      00:00:02   Te 0/1 (Up)
Dell#configure
Dell(conf)#interface port-channel 1
Dell(conf-if-po-1)#switchport backup interface port-channel 2
Apr 9 00:15:13: %STKUNIT0-M:CP %IFMGR-5-L2BKUP_WARN: Do not run any Layer2
protocols on Po 1 and Po 2
Apr 9 00:15:13: %STKUNIT0-M:CP %IFMGR-5-OSTATE_DN: Changed interface state to
down: Po 2
Apr 9 00:15:13: %STKUNIT0-M:CP %IFMGR-5-STATE_ACT_STBY: Changed interface state
to standby: Po 2
Dell(conf-if-po-1)#
Dell#
Dell#show interfaces switchport backup
Interface        Status   Paired Interface     Status
Port-channel 1   Active   Port-chato mannel 2  Standby
Port-channel 2   Standby  Port-channel 1       Active
Dell#

Dell(conf-if-po-1)#switchport backup interface tengigabitethernet 0/2
Apr 9 00:16:29: %STKUNIT0-M:CP %IFMGR-5-L2BKUP_WARN: Do not run any Layer2
protocols on Po 1 and Te 0/2
Dell(conf-if-po-1)#
```

Far-End Failure Detection

Far-end failure detection (FEFD) is supported on the S6000 platform.

FEFD is a protocol that senses remote data link errors in a network. FEFD responds by sending a unidirectional report that triggers an echoed response after a specified time interval. You can enable FEFD globally or locally on an interface basis. Disabling the global FEFD configuration does not disable the interface configuration.



```
Dell(conf-if-gi-4/0)#show config
!
interface TenGigabitEthernet 4/0
 no ip address
 switchport
 fefd
 no shutdown
```

```
Dell(conf-if-gi-1/0)#show config
!
interface TenGigabitEthernet 1/0
 no ip address
 switchport
 fefd
 no shutdown
```

```
2w0d4h : FEFD packet sent via interface Gi 1/0
   Sender state -- Bi-directional
   Sender info -- Mgmt Mac(00:01:e8:14:89:25), Slot-Port(Gi 1/0)
   Peer info -- Mgmt Mac (00:01:e8:14:89:25), Slot-Port(Gi 4/0)
   Sender hold time -- 3 (second)
```

```
2w0d4h : FEFD packet sent via interface Gi 4/0
   Sender state -- Bi-directional
   Sender info -- Mgmt Mac(00:01:e8:14:89:25), Slot-Port(Gi 4/0)
   Peer info -- Mgmt Mac (00:01:e8:14:89:25), Slot-Port(Gi 1/0)
   Sender hold time -- 3 (second)
```

Figure 65. Configuring Far-End Failure Detection

The report consists of several packets in SNAP format that are sent to the nearest known MAC address.

In the event of a far-end failure, the device stops receiving frames and, after the specified time interval, assumes that the far-end is not available. The connecting line protocol is brought down so that upper layer protocols can detect the neighbor unavailability faster.

FEFD State Changes

FEFD has two operational modes, Normal and Aggressive.

When you enable Normal mode on an interface and a far-end failure is detected, no intervention is required to reset the interface to bring it back to an FEFD operational state. When you enable Aggressive mode on an interface in the same state, manual intervention is required to reset the interface.

FEFD enabled systems (comprised of one or more interfaces) automatically switchs between four different states: Idle, Unknown, Bi-directional, and Err-disabled.

1. An interface on which FEFD is not configured is in Normal mode by default.
2. After you enable FEFD on an interface, it transitions to the Unknown state and sends an FEFD packet to the remote end of the link.
3. When the local interface receives the echoed packet from the remote end, the local interface transitions to the Bi-directional state.

4. If the FEFD enabled system is configured to use FEFD in Normal mode and neighboring echoes are not received after three intervals, (you can set each interval can be set between 3 and 300 seconds) the state changes to unknown.
5. If the FEFD system has been set to Aggressive mode and neighboring echoes are not received after three intervals, the state changes to Err-disabled. You must manually reset all interfaces in the Err-disabled state using the `fefd reset [interface]` command in EXEC privilege mode (it can be done globally or one interface at a time) before the FEFD enabled system can become operational again.

Table 36. State Change When Configuring FEFD

Local Event	Mode	Local State	Remote State	Local Admin Status	Local Protocol Status	Remote Admin Status	Remote Protocol Status
Shutdown	Normal	Admin Shutdown	Unknown	Down	Down	Up	Down
Shutdown	Aggressive	Admin Shutdown	Err-disabled	Up	Down	Up	Down
FEFD enable	Normal	Bi-directional	Bi-directional	Up	Up	Up	Up
FEFD enable	Aggressive	Bi-directional	Bi-directional	Up	Up	Up	Up
FEFD + FEFD disable	Normal	Locally disabled	Unknown	Up	Down	Up	Down
FEFD + FEFD disable	Aggressive	Locally disabled	Err-disabled	Up	Down	Up	Down
Link Failure	Normal	Unknown	Unknown	Up	Down	Up	Down
Link Failure	Aggressive	Err-disabled	Err-disabled	Up	Down	Up	Down

Important Points to Remember

- FEFD enabled ports are subject to an 8 to 10 second delay during an RPM failover before becoming operational.
- You can enable FEFD globally or on a per-interface basis. Interface FEFD configurations override global FEFD configurations.
- Dell Networking OS supports FEFD on physical Ethernet interfaces only, excluding the management interface.

Configuring FEFD

You can configure FEFD for all interfaces from CONFIGURATION mode, or on individual interfaces from INTERFACE mode.
To enable FEFD globally on all interfaces, use the following command.

- Enable FEFD globally on all interfaces.

  CONFIGURATION mode

  ```
  fefd-global
  ```

To report interval frequency and mode adjustments, use the following commands.

1. Setup two or more connected interfaces for Layer 2 or Layer 3.
   INTERFACE mode

   ```
   ip address ip address, switchport
   ```
2. Activate the necessary ports administratively.
   INTEFACE mode

   ```
   no shutdown
   ```
3. Enable fefd globally.
   CONFIGURATION mode

   ```
   fefd {interval | mode}
   ```

Example of the `show fefd` Command

To display information about the state of each interface, use the `show fefd` command in EXEC privilege mode.

```
Dell#show fefd
FEFD is globally 'ON', interval is 3 seconds, mode is 'Normal'.

INTERFACE   MODE    INTERVAL    STATE
                    (second)
Gi 1/0      Normal 3           Bi-directional
Gi 1/1      Normal 3           Admin Shutdown
Gi 1/2      Normal 3           Admin Shutdown
Gi 1/3      Normal 3           Admin Shutdown

Dell#show run fefd
!
fefd-global mode normal
fefd-global interval 3
```

Enabling FEFD on an Interface

To enable, change, or disable FEFD on an interface, use the following commands.

- Enable FEFD on a per interface basis.
  INTERFACE mode

  ```
  fefd
  ```
- Change the FEFD mode.
  INTERFACE mode

  ```
  fefd [mode {aggressive | normal}]
  ```
- Disable FEFD protocol on one interface.
  INTERFACE mode

  ```
  fefd disable
  ```

  Disabling an interface shuts down all protocols working on that interface’s connected line. It does not delete your previous FEFD configuration which you can enable again at any time.

To set up and activate two or more connected interfaces, use the following commands.

1. Setup two or more connected interfaces for Layer 2 or Layer 3.

   INTERFACE mode

   ```
   ip address ip address, switchport
   ```

2. Activate the necessary ports administratively.

   INTERFACE mode

   ```
   no shutdown
   ```

3. INTERFACE mode

   ```
   fefd {disable | interval | mode}
   ```

Example of Viewing FEFD Configuration

```
Dell(conf-if-gi-1/0)#show config
!
interface GigabitEthernet 1/0
  no ip address
  switchport
  fefd mode normal
  no shutdown

Dell(conf-if-gi-1/0)#do show fefd | grep 1/0
Gi 1/0          Normal        3           Unknown
```

Debugging FEFD

To debug FEFD, use the first command. To provide output for each packet transmission over the FEFD enabled connection, use the second command.

- Display output whenever events occur that initiate or disrupt an FEFD enabled connection.

  EXEC Privilege mode

  ```
  debug fefd events
  ```

- Provide output for each packet transmission over the FEFD enabled connection.

  EXEC Privilege mode

  ```
  debug fefd packets
  ```

Examples of the `debug fefd` Commands

```
Dell#debug fefd events
Dell#config
Dell(conf)#int gi 1/0
Dell(conf-if-gi-1/0)#shutdown
2w1d22h: %RPM0-P:CP %IFMGR-5-ASTATE_DN: Changed interface Admin state to down:
Gi 1/0
Dell(conf-if-gi-1/0)#2w1d22h : FEFD state on Gi 1/0 changed from ANY to Unknown
2w1d22h: %RPM0-P:CP %IFMGR-5-OSTATE_DN: Changed interface state to down: Gi 1/0
2w1d22h: %RPM0-P:CP %IFMGR-5-OSTATE_DN: Changed interface state to down: Gi 4/0
2w1d22h: %RPM0-P:CP %IFMGR-5-INACTIVE: Changed Vlan interface state to
inactive: Vl 1
2w1d22h : FEFD state on Gi 4/0 changed from Bi-directional to Unknown

Dell#debug fefd packets
Dell#2w1d22h : FEFD packet sent via interface Gi 1/0
```

```
   Sender state -- Bi-directional
   Sender info -- Mgmt Mac(00:01:e8:14:89:25), Slot-Port(Gi 1/0)
   Peer info -- Mgmt Mac (00:01:e8:14:89:25), Slot-Port(Gi 4/0)
   Sender hold time -- 3 (second)

2w1d22h : FEFD packet received on interface Gi 4/0
   Sender state -- Bi-directional
   Sender info -- Mgmt Mac(00:01:e8:14:89:25), Slot-Port(Gi 1/0)
   Peer info -- Mgmt Mac (00:01:e8:14:89:25), Slot-Port(Gi 4/0)
   Sender hold time -- 3 (second)
```

An RPM Failover

In the event that an RPM failover occurs, FEFD becomes operationally down on all enabled ports for approximately 8-10 seconds before automatically becoming operational again.

```
02-05-2009    12:40:38    Local7.Debug    10.16.151.12    Feb 5 07:06:09:
%RPM1-S:CP %RAM-6-FAILOVER_REQ: RPM failover request from active peer: User
request.
02-05-2009    12:40:38    Local7.Debug    10.16.151.12    Feb 5 07:06:19:
%RPM1-P:CP %IFMGR-5-OSTATE_UP: Changed interface state to up: Gi 0/45
02-05-2009    12:40:38    Local7.Debug    10.16.151.12    Feb 5 07:06:19:
%RPM1-P:CP %FEFD-5-FEFD-BIDIRECTION-LINK-DETECTED: Interface Gi 0/45 has
bidirectional link with its
peer
```

26

Link Layer Discovery Protocol (LLDP)

The link layer discovery protocol (LLDP) is supported on the S6000 platform.

802.1AB (LLDP) Overview

LLDP — defined by IEEE 802.1AB — is a protocol that enables a local area network (LAN) device to advertise its configuration and receive configuration information from adjacent LLDP-enabled LAN infrastructure devices.

The collected information is stored in a management information base (MIB) on each device, and is accessible via simple network management protocol (SNMP).

Protocol Data Units

Configuration information is exchanged in the form of Type, Length, Value (TLV) segments.

- Type — The kind of information included in the TLV.
- Length — The value, in octets, of the TLV after the Length field.
- Value — The configuration information that the agent is advertising.

The chassis ID TLV is shown in the following illustration.

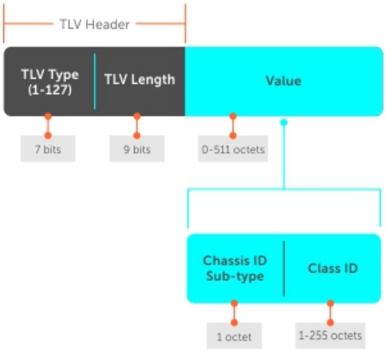


Figure 66. Type, Length, Value (TLV) Segment

TLVs are encapsulated in a frame called an LLDP data unit (LLDPDU) (shown in the following table), which is transmitted from one LLDP-enabled device to its LLDP-enabled neighbors. LLDP is a one-way protocol. LLDP-enabled devices (LLDP agents) can transmit and/or receive advertisements, but they cannot solicit and do not respond to advertisements.

There are five types of TLVs. All types are mandatory in the construction of an LLDPDU except Optional TLVs. You can configure the inclusion of individual Optional TLVs.

Table 37. Type, Length, Value (TLV) Types

Type	TLV	Description
0	End of LLDPDU	Marks the end of an LLDPDU.
1	Chassis ID	An administratively assigned name that identifies the LLDP agent.
2	Port ID	An administratively assigned name that identifies a port through which TLVs are sent and received.
3	Time to Live	An administratively assigned name that identifies a port through which TLVs are sent and received.
—	Optional	Includes sub-types of TLVs that advertise specific configuration information. These sub-types are Management TLVs, IEEE 802.1, IEEE 802.3, and TIA-1057 Organizationally Specific TLVs.

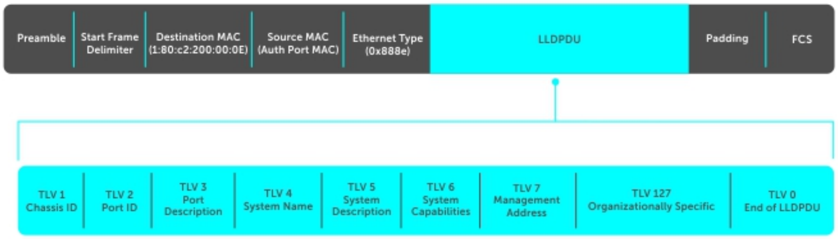

Preamble
Start Frame Delimiter
Destination MAC (1:80:c2:200:00:0E)
Source MAC (Auth Port MAC)
Ethernet Type (0x888e)
LLDPDU
Padding
FCS
TLV 1 Chassis ID
TLV 2 Port ID
TLV 3 Port Description
TLV 4 System Name
TLV 5 System Description
TLV 6 System Capabilities
TLV 7 Management Address
TLV 127 Organizationally Specific
TLV 0 End of LLDPDU

Figure 67. LLDPDU Frame

Optional TLVs

The Dell Networking OS supports these optional TLVs: management TLVs, IEEE 802.1 and 802.3 organizationally specific TLVs, and TIA-1057 organizationally specific TLVs.

Management TLVs

A management TLV is an optional TLVs sub-type. This kind of TLV contains essential management information about the sender.

Organizationally Specific TLVs

A professional organization or a vendor can define organizationally specific TLVs. They have two mandatory fields (as shown in the following illustration) in addition to the basic TLV fields.

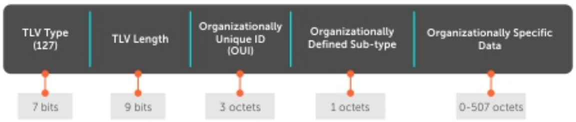

TLV Type (127)
TLV Length
Organizationally Unique ID (OUI)
Organizationally Defined Sub-type
Organizationally Specific Data
7 bits
9 bits
3 octets
1 octets
0-507 octets

Figure 68. Organizationally Specific TLV

IEEE Organizationally Specific TLVs

Eight TLV types have been defined by the IEEE 802.1 and 802.3 working groups as a basic part of LLDP; the IEEE OUI is 00-80-C2. You can configure the Dell Networking system to advertise any or all of these TLVs.

Table 38. Optional TLV Types

Type	TLV	Description
Optional TLVs		
4	Port description	A user-defined alphanumeric string that describes the port. Dell Networking OS does not currently support this TLV.
5	System name	A user-defined alphanumeric string that identifies the system.
6	System description	A user-defined alphanumeric string that identifies the system.
7	System capabilities	Identifies the chassis as one or more of the following: repeater, bridge, WLAN Access Point, Router, Telephone, DOCSIS cable device, end station only, or other.
8	Management address	Indicates the network address of the management interface. Dell Networking OS does not currently support this TLV.
IEEE 802.1 Organizationally Specific TLVs		
127	Port-VLAN ID	On Dell Networking systems, indicates the untagged VLAN to which a port belongs.
127	Port and Protocol VLAN ID	On Dell Networking systems, indicates the tagged VLAN to which a port belongs (and the untagged VLAN to which a port belongs if the port is in Hybrid mode).

Type	TLV	Description
127	Protocol Identity	Indicates the protocols that the port can process. Dell Networking OS does not currently support this TLV.
IEEE 802.3 Organizationally Specific TLVs		
127	MAC/PHY Configuration/Status	Indicates the capability and current setting of the duplex status and bit rate, and whether the current settings are the result of auto-negotiation. This TLV is not available in the Dell Networking OS implementation of LLDP, but is available and mandatory (non-configurable) in the LLDP-MED implementation.
127	Power via MDI	Dell Networking supports the LLDP-MED protocol, which recommends that Power via MDI TLV be not implemented, and therefore Dell Networking implements Extended Power via MDI TLV only.
127	Link Aggregation	Indicates whether the link is capable of being aggregated, whether it is currently in a LAG, and the port identification of the LAG. Dell Networking OS does not currently support this TLV.
127	Maximum Frame Size	Indicates the maximum frame size capability of the MAC and PHY.

TIA-1057 (LLDP-MED) Overview

Link layer discovery protocol — media endpoint discovery (LLDP-MED) as defined by ANSI/ TIA-1057— provides additional organizationally specific TLVs so that endpoint devices and network connectivity devices can advertise their characteristics and configuration information; the OUI for the Telecommunications Industry Association (TIA) is 00-12-BB.

- **LLDP-MED Endpoint Device** — any device that is on an IEEE 802 LAN network edge can communicate using IP and uses the LLDP-MED framework.
- **LLDP-MED Network Connectivity Device** — any device that provides access to an IEEE 802 LAN to an LLDP-MED endpoint device and supports IEEE 802.1AB (LLDP) and TIA-1057 (LLDP-MED). The Dell Networking system is an LLDP-MED network connectivity device.

Regarding connected endpoint devices, LLDP-MED provides network connectivity devices with the ability to:

- manage inventory
- manage Power over Ethernet (PoE)
- identify physical location
- identify network policy

LLDP-MED is designed for, but not limited to, VoIP endpoints.

TIA Organizationally Specific TLVs

The Dell Networking system is an LLDP-MED Network Connectivity Device (Device Type 4).

Network connectivity devices are responsible for:

- transmitting an LLDP-MED capability TLV to endpoint devices
- storing the information that endpoint devices advertise

The following table describes the five types of TIA-1057 Organizationally Specific TLVs.

Table 39. TIA-1057 (LLDP-MED) Organizationally Specific TLVs

Type	SubType	TLV	Description
127	1	LLDP-MED Capabilities	Indicates: • whether the transmitting device supports LLDP-MED • what LLDP-MED TLVs it supports • LLDP device class
127	2	Network Policy	Indicates the application type, VLAN ID, Layer 2 Priority, and DSCP value.
127	3	Location Identification	Indicates that the physical location of the device expressed in one of three possible formats: • Coordinate Based LCI • Civic Address LCI • Emergency Call Services ELIN
127	4	Location Identification	Indicates power requirements, priority, and power status.
Inventory Management TLVs	Implementation of this set of TLVs is optional in LLDP-MED devices.		

Type	SubType	TLV	Description
	None or all TLVs must be supported. Dell Networking OS does not currently support these TLVs.		
127	5	Inventory — Hardware Revision	Indicates the hardware revision of the LLDP-MED device.
127	6	Inventory — Firmware Revision	Indicates the firmware revision of the LLDP-MED device.
127	7	Inventory — Software Revision	Indicates the software revision of the LLDP-MED device.
127	8	Inventory — Serial Number	Indicates the device serial number of the LLDP-MED device.
127	9	Inventory — Manufacturer Name	Indicates the manufacturer of the LLDP-MED device.
127	10	Inventory — Model Name	Indicates the model of the LLDP-MED device.
127	11	Inventory — Asset ID	Indicates a user specified device number to manage inventory.
127	12–255	Reserved	—

LLDP-MED Capabilities TLV

The LLDP-MED capabilities TLV communicates the types of TLVs that the endpoint device and the network connectivity device support. LLDP-MED network connectivity devices must transmit the Network Policies TLV.

- The value of the LLDP-MED capabilities field in the TLV is a 2–octet bitmap, each bit represents an LLDP-MED capability (as shown in the following table).
- The possible values of the LLDP-MED device type are shown in the following. The Dell Networking system is a network connectivity device, which is Type 4.

When you enable LLDP-MED in Dell Networking OS (using the `advertise med` command), the system begins transmitting this TLV.

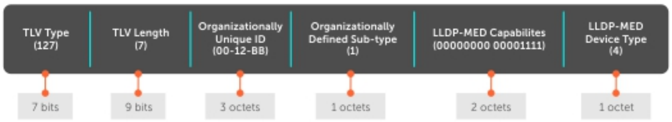

TLV Type
(127)
TLV Length
(7)
Organizationally
Unique ID
(00-12-BB)
Organizationally
Defined Sub-type
(1)
LLDP-MED Capabilites
(00000000 00001111)
LLDP-MED
Device Type
(4)
7 bits
9 bits
3 octets
1 octets
2 octets
1 octet

Figure 69. LLDP-MED Capabilities TLV

Table 40. Dell Networking OS LLDP-MED Capabilities

Bit Position	TLV	Dell Networking OS Support
0	LLDP-MED Capabilities	Yes
1	Network Policy	Yes
2	Location Identification	Yes
3	Extended Power via MDI-PSE	Yes
4	Extended Power via MDI-PD	No
5	Inventory	No
6–15	reserved	No

Table 41. LLDP-MED Device Types

Value	Device Type
0	Type Not Defined
1	Endpoint Class 1
2	Endpoint Class 2
3	Endpoint Class 3
4	Network Connectivity
5–255	Reserved

LLDP-MED Network Policies TLV

A network policy in the context of LLDP-MED is a device's VLAN configuration and associated Layer 2 and Layer 3 configurations.

LLDP-MED network policies TLV include:

- VLAN ID
- VLAN tagged or untagged status
- Layer 2 priority
- DSCP value

An integer represents the application type (the Type integer shown in the following table), which indicates a device function for which a unique network policy is defined. An individual LLDP-MED network policy TLV is generated for each application type that you specify with the Dell Networking OS CLI (Advertising TLVs).

NOTE: As shown in the following table, signaling is a series of control packets that are exchanged between an endpoint device and a network connectivity device to establish and maintain a connection. These signal packets might require a different network policy than the media packets for which a connection is made. In this case, configure the signaling application.

Table 42. Network Policy Applications

Type	Application	Description
0	Reserved	—
1	Voice	Specify this application type for dedicated IP telephony handsets and other appliances supporting interactive voice services.
2	Voice Signaling	Specify this application type only if voice control packets use a separate network policy than voice data.
3	Guest Voice	Specify this application type for a separate limited voice service for guest users with their own IP telephony handsets and other appliances supporting interactive voice services.
4	Guest Voice Signaling	Specify this application type only if guest voice control packets use a separate network policy than voice data.
5	Softphone Voice	Specify this application type only if guest voice control packets use a separate network policy than voice data.
6	Video Conferencing	Specify this application type for dedicated video conferencing and other similar appliances supporting real-time interactive video.
7	Streaming Video	Specify this application type for dedicated video conferencing and other similar appliances supporting real-time interactive video.
8	Video Signaling	Specify this application type only if video control packets use a separate network policy than video data.
9–255	Reserved	—

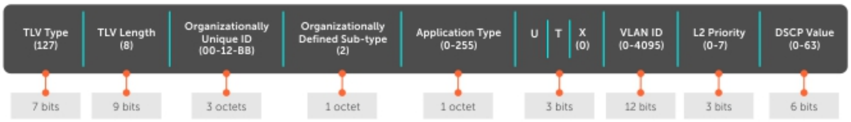

TLV Type (127)
TLV Length (8)
Organizationally Unique ID (00-12-BB)
Organizationally Defined Sub-type (2)
Application Type (0-255)
U
T
X (0)
VLAN ID (0-4095)
L2 Priority (0-7)
DSCP Value (0-63)
7 bits
9 bits
3 octets
1 octet
1 octet
3 bits
12 bits
3 bits
6 bits

Figure 70. LLDP-MED Policies TLV

Extended Power via MDI TLV

The extended power via MDI TLV enables advanced PoE management between LLDP-MED endpoints and network connectivity devices.

Advertise the extended power via MDI on all ports that are connected to an 802.3af powered, LLDP-MED endpoint device.

- **Power Type** — there are two possible power types: power source entity (PSE) or power device (PD). The Dell Networking system is a PSE, which corresponds to a value of 0, based on the TIA-1057 specification.
- **Power Source** — there are two possible power sources: primary and backup. The Dell Networking system is a primary power source, which corresponds to a value of 1, based on the TIA-1057 specification.
- **Power Priority** — there are three possible priorities: Low, High, and Critical. On Dell Networking systems, the default power priority is **High**, which corresponds to a value of 2 based on the TIA-1057 specification. You can configure a different power priority through the CLI. Dell Networking also honors the power priority value the powered device sends; however, the CLI configuration takes precedence.
- **Power Value** — Dell Networking advertises the maximum amount of power that can be supplied on the port. By default the power is **15.4W**, which corresponds to a power value of 130, based on the TIA-1057 specification. You can advertise a different power value using the `max-milliwatts` option with the `power inline auto | static` command. Dell Networking also honors the power value (power requirement) the powered device sends when the port is configured for `power inline auto`.

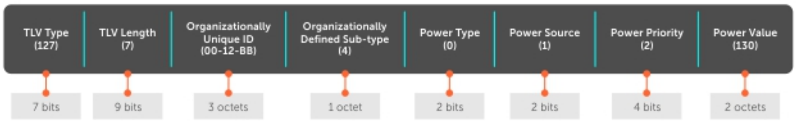


Figure 71. Extended Power via MDI TLV

Configure LLDP

Configuring LLDP is a two-step process.

1. Enable LLDP globally.
2. Advertise TLVs out of an interface.

Related Configuration Tasks

- Viewing the LLDP Configuration
- Viewing Information Advertised by Adjacent LLDP Agents
- Configuring LLDPDU Intervals
- Configuring Transmit and Receive Mode
- Configuring a Time to Live
- Debugging LLDP

Important Points to Remember

- LLDP is enabled by default.
- Dell Networking systems support up to eight neighbors per interface.
- Dell Networking systems support a maximum of 8000 total neighbors per system. If the number of interfaces multiplied by eight exceeds the maximum, the system does not configure more than 8000.
- INTERFACE level configurations override all CONFIGURATION level configurations.
- LLDP is not hitless.

LLDP Compatibility

- Spanning tree and force10 ring protocol “blocked” ports allow LLDPDUs.
- 802.1X controlled ports do not allow LLDPDUs until the connected device is authenticated.

CONFIGURATION versus INTERFACE Configurations

All LLDP configuration commands are available in PROTOCOL LLDP mode, which is a sub-mode of the CONFIGURATION mode and INTERFACE mode.

- Configurations made at the CONFIGURATION level are global; that is, they affect all interfaces on the system.
- Configurations made at the INTERFACE level affect only the specific interface; they override CONFIGURATION level configurations.

Example of the `protocol lldp` Command (CONFIGURATION Level)

```
R1(conf)#protocol lldp
R1(conf-lldp)#?
advertise        Advertise TLVs
disable          Disable LLDP protocol globally
end              Exit from configuration mode
exit             Exit from LLDP configuration mode
hello            LLDP hello configuration
mode             LLDP mode configuration (default = rx and tx)
multiplier       LLDP multiplier configuration
no               Negate a command or set its defaults
show             Show LLDP configuration


Dell(conf-lldp)#exit
Dell(conf)#interface tengigabitethernet 0/3
Dell(conf-if-te-0/3)#protocol lldp
Dell(conf-if-te-0/3-lldp)#?
advertise        Advertise TLVs
disable          Disable LLDP protocol on this interface
end              Exit from configuration mode
exit             Exit from LLDP configuration mode
hello            LLDP hello configuration
mode             LLDP mode configuration (default = rx and tx)
multiplier       LLDP multiplier configuration
no               Negate a command or set its defaults
show             Show LLDP configuration
Dell(conf-if-te-0/3-lldp)#
```

Enabling LLDP

LLDP is enabled by default. Enable and disable LLDP globally or per interface. If you enable LLDP globally, all UP interfaces send periodic LLDPDUs.
To enable LLDP, use the following command.

1. Enter Protocol LLDP mode.
   CONFIGURATION or INTERFACE mode

   ```
   protocol lldp
   ```

2. Enable LLDP.
   PROTOCOL LLDP mode

   ```
   no disable
   ```

Disabling and Undoing LLDP

To disable or undo LLDP, use the following command.

- Disable LLDP globally or for an interface.
  `disable`

To undo an LLDP configuration, precede the relevant command with the keyword `no`.

Enabling LLDP on Management Ports

LLDP on management ports is enabled by default.
To enable LLDP on management ports, use the following command.

1. Enter Protocol LLDP mode.
   CONFIGURATION mode

   ```
   protocol lldp
   ```

2. Enable LLDP.
   PROTOCOL LLDP mode

   ```
   no disable
   ```

Disabling and Undoing LLDP on Management Ports

To disable or undo LLDP on management ports, use the following command.

1. Enter Protocol LLDP mode.
   CONFIGURATION mode.

   ```
   protocol lldp
   ```

2. Enter LLDP management-interface mode.
   LLDP-MANAGEMENT-INTERFACE mode.

   ```
   management-interface
   ```

3. Enter the `disable` command.

   LLDP-MANAGEMENT-INTERFACE mode.

To undo an LLDP management port configuration, precede the relevant command with the keyword `no`.

Advertising TLVs

You can configure the system to advertise TLVs out of all interfaces or out of specific interfaces.

- If you configure the system globally, all interfaces send LLDPDUs with the specified TLVs.
- If you configure an interface, only the interface sends LLDPDUs with the specified TLVs.
- If you configure LLDP both globally and at interface level, the interface level configuration overrides the global configuration.

To advertise TLVs, use the following commands.

1. Enter LLDP mode.

   CONFIGURATION or INTERFACE mode

   ```
   protocol lldp
   ```
2. Advertise one or more TLVs.

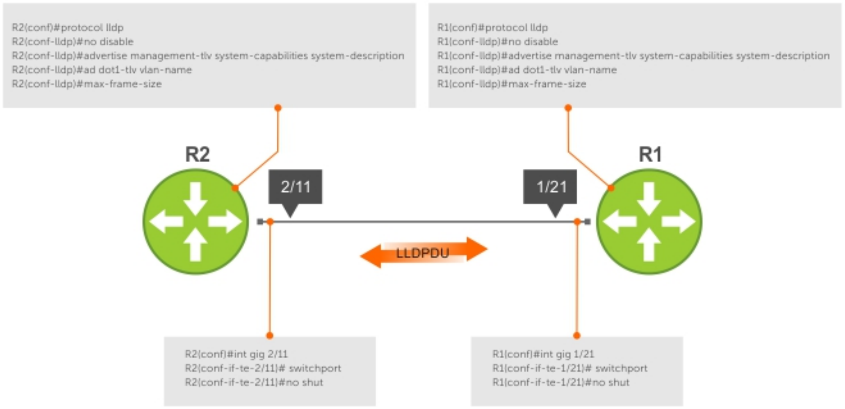
   PROTOCOL LLDP mode

   ```
   advertise {dcbx-appln-tlv | dcbx-tlv | dot3-tlv | interface-port-desc |
   management-tlv | med }
   ```

   Include the keyword for each TLV you want to advertise.

   - For management TLVs: `system-capabilities, system-description`.
   - For 802.1 TLVs: `port-protocol-vlan-id, port-vlan-id` .
   - For 802.3 TLVs: `max-frame-size`.
   - For TIA-1057 TLVs:
     - `guest-voice`
     - `guest-voice-signaling`
     - `location-identification`
     - `power-via-mdi`
     - `softphone-voice`
     - `streaming-video`
     - `video-conferencing`
     - `video-signaling`
     - `voice`
     - `voice-signaling`

In the following example, LLDP is enabled globally. R1 and R2 are transmitting periodic LLDPDUs that contain management, 802.1, and 802.3 TLVs.



```
R2(conf)#protocol lldp
R2(conf-lldp)#no disable
R2(conf-lldp)#advertise management-tlv system-capabilities system-description
R2(conf-lldp)#ad dot1-tlv vlan-name
R2(conf-lldp)#max-frame-size
```

```
R1(conf)#protocol lldp
R1(conf-lldp)#no disable
R1(conf-lldp)#advertise management-tlv system-capabilities system-description
R1(conf-lldp)#ad dot1-tlv vlan-name
R1(conf-lldp)#max-frame-size
```

```
R2(conf)#int gig 2/11
R2(conf-if-te-2/11)# switchport
R2(conf-if-te-2/11)#no shut
```

```
R1(conf)#int gig 1/21
R1(conf-if-te-1/21)# switchport
R1(conf-if-te-1/21)#no shut
```

Figure 72. Configuring LLDP

Viewing the LLDP Configuration

To view the LLDP configuration, use the following command.

- Display the LLDP configuration.

  CONFIGURATION or INTERFACE mode

  ```
  show config
  ```

Examples of Viewing LLDP Configurations

```
Dell(conf)#protocol lldp
Dell(conf-lldp)#show config
!
protocol lldp
  advertise dot1-tlv port-protocol-vlan-id port-vlan-id
  advertise dot3-tlv max-frame-size
  advertise management-tlv system-capabilities system-description
  hello 10
  no disable
Dell(conf-lldp)#

Dell(conf-lldp)#exit
Dell(conf)#interface gigabitethernet 1/31
Dell(conf-if-gi-1/31)#show config
!
interface GigabitEthernet 1/31
  no ip address
  switchport
  no shutdown
Dell(conf-if-gi-1/31)#protocol lldp
Dell(conf-if-gi-1/31-lldp)#show config
!
  protocol lldp
Dell(conf-if-gi-1/31-lldp)#
```

Viewing Information Advertised by Adjacent LLDP Agents

To view brief information about adjacent devices or to view all the information that neighbors are advertising, use the following commands.

- Display brief information about adjacent devices.

  `show lldp neighbors`
- Display all of the information that neighbors are advertising.

  `show lldp neighbors detail`

Example of Viewing Brief Information Advertised by Neighbors

```
Dell(conf-if-te-0/3-lldp)#end
Dell (conf-if-te-0/3)#do show lldp neighbors
Loc PortID  Rem Host Name  Rem Port Id           Rem Chassis Id
--------------------------------------------------------------------
Te 0/1     -           TenGigabitEthernet 0/5  00:01:e8:05:40:46
Te 0/2     -           TenGigabitEthernet 0/6  00:01:e8:05:40:46
Dell (conf-if-te-0/3)#
```

Example of Viewing Details Advertised by Neighbors

```
Dell#show lldp neighbors detail
========================================================================
Local Interface Te 0/4 has 1 neighbor
  Total Frames Out: 6547
  Total Frames In: 4136
  Total Neighbor information Age outs: 0
  Total Frames Discarded: 0
  Total In Error Frames: 0
  Total Unrecognized TLVs: 0
  Total TLVs Discarded: 0
  Next packet will be sent after 7 seconds
  The neighbors are given below:
-----------------------------------------------------------------------

  Remote Chassis ID Subtype: Mac address (4)
  Remote Chassis ID: 00:01:e8:06:95:3e
  Remote Port Subtype: Interface name (5)
  Remote Port ID: TeGigabitEthernet 2/11
  Local Port ID: TeGigabitEthernet 1/21
  Locally assigned remote Neighbor Index: 4
  Remote TTL: 120
  Information valid for next 120 seconds
  Time since last information change of this neighbor: 01:50:16
  Remote MTU: 1554
  Remote System Desc: Dell Networks Real Time Operating System Software
  Dell Operating System Version: 1.0.
  Dell Application Software Version: 9.4.0.0. Copyright (c) 1999-2014
  Build Time: Thu Aug 9 01:05:51 PDT 1999-2014
  Existing System Capabilities: Repeater Bridge Router
  Enabled System Capabilities: Repeater Bridge Router
  Remote Port Vlan ID: 1
  Port and Protocol Vlan ID: 1, Capability: Supported, Status: Enabled
  -------------------------------------------------------------------------

========================================================================
```

Configuring LLDPDU Intervals

LLDPDUs are transmitted periodically; the default interval is **30 seconds**.

To configure LLDPDU intervals, use the following command.

- Configure a non-default transmit interval.

  CONFIGURATION mode or INTERFACE mode

  ```
  hello
  ```

Example of Viewing LLDPDU Intervals

```
R1(conf)#protocol lldp
R1(conf-lldp)#show config
!
protocol lldp
  advertise dot1-tlv port-protocol-vlan-id port-vlan-id
  advertise dot3-tlv max-frame-size
  advertise management-tlv system-capabilities system-description
  no disable
R1(conf-lldp)#mode ?
rx          Rx only
tx          Tx only
R1(conf-lldp)#mode tx
R1(conf-lldp)#show config
!
protocol lldp
  advertise dot1-tlv port-protocol-vlan-id port-vlan-id
  advertise dot3-tlv max-frame-size
  advertise management-tlv system-capabilities system-description
  mode tx
  no disable
R1(conf-lldp)#no mode
R1(conf-lldp)#show config
!
protocol lldp
  advertise dot1-tlv port-protocol-vlan-id port-vlan-id
  advertise dot3-tlv max-frame-size
  advertise management-tlv system-capabilities system-description
  no disable
R1(conf-lldp)#
```

Configuring Transmit and Receive Mode

After you enable LLDP, Dell Networking systems transmit *and* receive LLDPDUs by default.
To configure the system to transmit or receive only and return to the default, use the following commands.

- Transmit only.

  CONFIGURATION mode or INTERFACE mode

  ```
  mode tx
  ```
- Receive only.

  CONFIGURATION mode or INTERFACE mode

  ```
  mode rx
  ```

- Return to the default setting.

  CONFIGURATION mode or INTERFACE mode

  ```
  no mode
  ```

Example of Configuring a Single Mode

```
R1(conf)#protocol lldp
R1(conf-lldp)#show config
!
protocol lldp
  advertise dot1-tlv port-protocol-vlan-id port-vlan-id
  advertise dot3-tlv max-frame-size
  advertise management-tlv system-capabilities system-description
  no disable
R1(conf-lldp)#mode ?
rx              Rx only
tx              Tx only
R1(conf-lldp)#mode tx
R1(conf-lldp)#show config
!
protocol lldp
  advertise dot1-tlv port-protocol-vlan-id port-vlan-id
  advertise dot3-tlv max-frame-size
  advertise management-tlv system-capabilities system-description
  mode tx
  no disable
R1(conf-lldp)#no mode
R1(conf-lldp)#show config
!
protocol lldp
  advertise dot1-tlv port-protocol-vlan-id port-vlan-id
  advertise dot3-tlv max-frame-size
  advertise management-tlv system-capabilities system-description
  no disable
R1(conf-lldp)#
```

Configuring a Time to Live

The information received from a neighbor expires after a specific amount of time (measured in seconds) called a time to live (TTL).
The TTL is the product of the LLDPDU transmit interval (hello) and an integer called a multiplier. The default multiplier is **4**, which results in a default TTL of 120 seconds.

- Adjust the TTL value.

  CONFIGURATION mode or INTERFACE mode.

  ```
  multiplier
  ```

- Return to the default multiplier value.

  CONFIGURATION mode or INTERFACE mode.

  ```
  no multiplier
  ```

Example of the `multiplier` Command to Configure Time to Live

```
R1(conf-lldp)#show config
!
protocol lldp
```

```
  advertise dot1-tlv port-protocol-vlan-id port-vlan-id
  advertise dot3-tlv max-frame-size
  advertise management-tlv system-capabilities system-description
  no disable
R1(conf-lldp)#multiplier ?
<2-10>            Multiplier (default=4)
R1(conf-lldp)#multiplier 5
R1(conf-lldp)#show config
!
protocol lldp
  advertise dot1-tlv port-protocol-vlan-id port-vlan-id
  advertise dot3-tlv max-frame-size
  advertise management-tlv system-capabilities system-description
  multiplier 5
  no disable
R1(conf-lldp)#no multiplier
R1(conf-lldp)#show config
!
protocol lldp
  advertise dot1-tlv port-protocol-vlan-id port-vlan-id
  advertise dot3-tlv max-frame-size
  advertise management-tlv system-capabilities system-description
  no disable
R1(conf-lldp)#
```

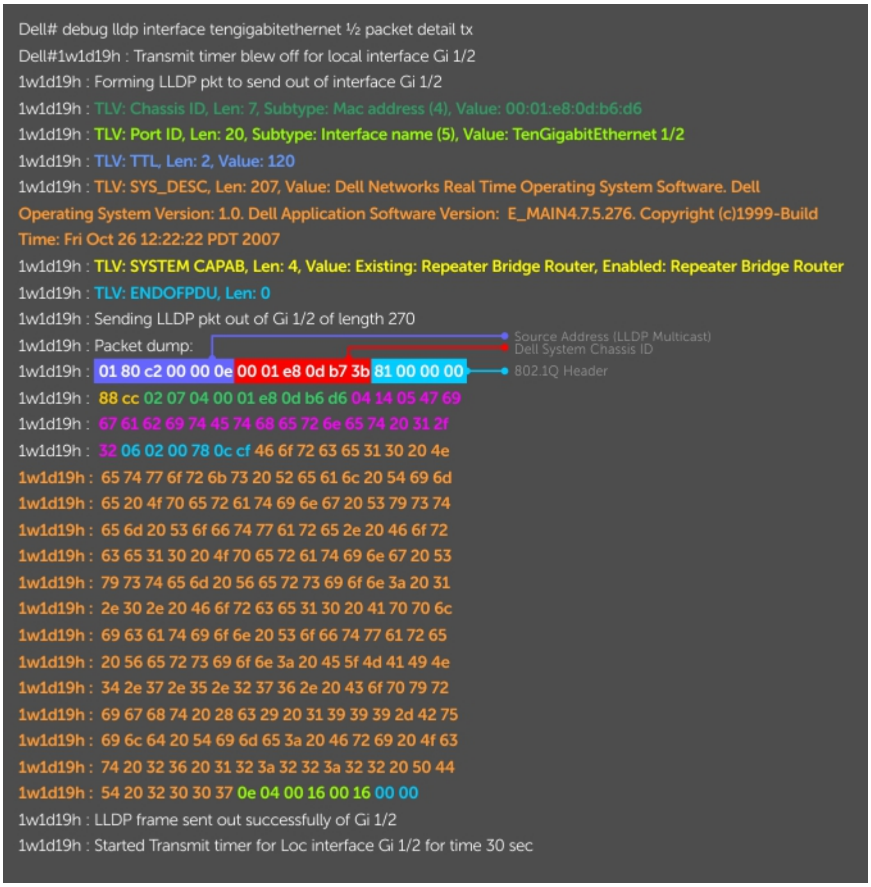
Debugging LLDP

You can view the TLVs that your system is sending and receiving.
To view the TLVs, use the following commands.

- View a readable version of the TLVs.

  `debug lldp brief`
- View a readable version of the TLVs plus a hexadecimal version of the entire LLDPDU.

  `debug lldp detail`



```
Dell# debug lldp interface tengigabitethernet ½ packet detail tx
Dell#1w1d19h : Transmit timer blew off for local interface Gi 1/2
1w1d19h : Forming LLDP pkt to send out of interface Gi 1/2
1w1d19h : TLV: Chassis ID, Len: 7, Subtype: Mac address (4), Value: 00:01:e8:0d:b6:d6
1w1d19h : TLV: Port ID, Len: 20, Subtype: Interface name (5), Value: TenGigabitEthernet 1/2
1w1d19h : TLV: TTL, Len: 2, Value: 120
1w1d19h : TLV: SYS_DESC, Len: 207, Value: Dell Networks Real Time Operating System Software. Dell
Operating System Version: 1.0. Dell Application Software Version:  E_MAIN4.7.5.276. Copyright (c)1999-Build
Time: Fri Oct 26 12:22:22 PDT 2007
1w1d19h : TLV: SYSTEM CAPAB, Len: 4, Value: Existing: Repeater Bridge Router, Enabled: Repeater Bridge Router
1w1d19h : TLV: ENDOFPDU, Len: 0
1w1d19h : Sending LLDP pkt out of Gi 1/2 of length 270
1w1d19h : Packet dump:
1w1d19h : 01 80 c2 00 00 0e 00 01 e8 0d b7 3b 81 00 00 00
1w1d19h :  88 cc 02 07 04 00 01 e8 0d b6 d6 04 14 05 47 69
1w1d19h :  67 61 62 69 74 45 74 68 65 72 6e 65 74 20 31 2f
1w1d19h :  32 06 02 00 78 0c cf 46 6f 72 63 65 31 30 20 4e
1w1d19h :  65 74 77 6f 72 6b 73 20 52 65 61 6c 20 54 69 6d
1w1d19h :  65 20 4f 70 65 72 61 74 69 6e 67 20 53 79 73 74
1w1d19h :  65 6d 20 53 6f 66 74 77 61 72 65 2e 20 46 6f 72
1w1d19h :  63 65 31 30 20 4f 70 65 72 61 74 69 6e 67 20 53
1w1d19h :  79 73 74 65 6d 20 56 65 72 73 69 6f 6e 3a 20 31
1w1d19h :  2e 30 2e 20 46 6f 72 63 65 31 30 20 41 70 70 6c
1w1d19h :  69 63 61 74 69 6f 6e 20 53 6f 66 74 77 61 72 65
1w1d19h :  20 56 65 72 73 69 6f 6e 3a 20 45 5f 4d 41 49 4e
1w1d19h :  34 2e 37 2e 35 2e 32 37 36 2e 20 43 6f 70 79 72
1w1d19h :  69 67 68 74 20 28 63 29 20 31 39 39 39 2d 42 75
1w1d19h :  69 6c 64 20 54 69 6d 65 3a 20 46 72 69 20 4f 63
1w1d19h :  74 20 32 36 20 31 32 3a 32 32 3a 32 32 20 50 44
1w1d19h :  54 20 32 30 30 37 0e 04 00 16 00 16 00 00
1w1d19h : LLDP frame sent out successfully of Gi 1/2
1w1d19h : Started Transmit timer for Loc interface Gi 1/2 for time 30 sec
```

Figure 73. The debug lldp detail Command — LLDPDU Packet Dissection

Relevant Management Objects

Dell Networking OS supports all IEEE 802.1AB MIB objects.

The following tables list the objects associated with:

- received and transmitted TLVs
- the LLDP configuration on the local agent
- IEEE 802.1AB Organizationally Specific TLVs
- received and transmitted LLDP-MED TLVs

Table 43. LLDP Configuration MIB Objects

MIB Object Category	LLDP Variable	LLDP MIB Object	Description
LLDP Configuration	adminStatus	lldpPortConfigAdminStatus	Whether you enable the local LLDP agent for transmit, receive, or both.
	msgTxHold	lldpMessageTxHoldMultiplier	Multiplier value.

MIB Object Category	LLDP Variable	LLDP MIB Object	Description
	msgTxInterval	lldpMessageTxInterval	Transmit Interval value.
	rxInfoTTL	lldpRxInfoTTL	Time to live for received TLVs.
	txInfoTTL	lldpTxInfoTTL	Time to live for transmitted TLVs.
Basic TLV Selection	mibBasicTLVsTxEnable	lldpPortConfigTLVsTxEnable	Indicates which management TLVs are enabled for system ports.
	mibMgmtAddrInstanceTxEnable	lldpManAddrPortsTxEnable	The management addresses defined for the system and the ports through which they are enabled for transmission.
LLDP Statistics	statsAgeoutsTotal	lldpStatsRxPortAgeoutsTotal	Total number of times that a neighbor's information is deleted on the local system due to an rxInfoTTL timer expiration.
	statsFramesDiscardedTotal	lldpStatsRxPortFramesDiscardedTotal	Total number of LLDP frames received then discarded.
	statsFramesInErrorsTotal	lldpStatsRxPortFramesErrors	Total number of LLDP frames received on a port with errors.
	statsFramesInTotal	lldpStatsRxPortFramesTotal	Total number of LLDP frames received through the port.
	statsFramesOutTotal	lldpStatsTxPortFramesTotal	Total number of LLDP frames transmitted through the port.
	statsTLVsDiscardedTotal	lldpStatsRxPortTLVsDiscardedTotal	Total number of TLVs received then discarded.
	statsTLVsUnrecognizedTotal	lldpStatsRxPortTLVsUnrecognizedTotal	Total number of all TLVs the local agent does not recognize.

Table 44. LLDP System MIB Objects

TLV Type	TLV Name	TLV Variable	System	LLDP MIB Object
1	Chassis ID	chassis ID subtype	Local	lldpLocChassisIdSubtype
			Remote	lldpRemChassisIdSubtype
		chassid ID	Local	lldpLocChassisId
			Remote	lldpRemChassisId
2	Port ID	port subtype	Local	lldpLocPortIdSubtype
			Remote	lldpRemPortIdSubtype
		port ID	Local	lldpLocPortId
			Remote	lldpRemPortId
4	Port Description	port description	Local	lldpLocPortDesc
			Remote	lldpRemPortDesc
5	System Name	system name	Local	lldpLocSysName
			Remote	lldpRemSysName
6	System Description	system description	Local	lldpLocSysDesc
			Remote	lldpRemSysDesc
7	System Capabilities	system capabilities	Local	lldpLocSysCapSupported
			Remote	lldpRemSysCapSupported
8	Management Address	enabled capabilities	Local	lldpLocSysCapEnabled
			Remote	lldpRemSysCapEnabled
		management address length	Local	lldpLocManAddrLen
			Remote	lldpRemManAddrLen
		management address subtype	Local	lldpLocManAddrSubtype
			Remote	lldpRemManAddrSubtype
		management address	Local	lldpLocManAddr
			Remote	lldpRemManAddr

TLV Type	TLV Name	TLV Variable	System	LLDP MIB Object
		interface numbering subtype	Local	lldpLocManAddrIfSubtype
			Remote	lldpRemManAddrIfSubtype
		interface number	Local	lldpLocManAddrIfId
			Remote	lldpRemManAddrIfId
		OID	Local	lldpLocManAddrOID
			Remote	lldpRemManAddrOID

Table 45. LLDP 802.1 Organizationally specific TLV MIB Objects

TLV Type	TLV Name	TLV Variable	System	LLDP MIB Object
127	Port-VLAN ID	PVID	Local	lldpXdot1LocPortVlanId
			Remote	lldpXdot1RemPortVlanId
127	Port and Protocol VLAN ID	port and protocol VLAN supported	Local	lldpXdot1LocProtoVlanSupported
			Remote	lldpXdot1RemProtoVlanSupported
		port and protocol VLAN enabled	Local	lldpXdot1LocProtoVlanEnabled
			Remote	lldpXdot1RemProtoVlanEnabled
		PPVID	Local	lldpXdot1LocProtoVlanId
			Remote	lldpXdot1RemProtoVlanId
127	VLAN Name	VID	Local	lldpXdot1LocVlanId
			Remote	lldpXdot1RemVlanId
		VLAN name length	Local	lldpXdot1LocVlanName
			Remote	lldpXdot1RemVlanName
		VLAN name	Local	lldpXdot1LocVlanName
			Remote	lldpXdot1RemVlanName

Table 46. LLDP-MED System MIB Objects

TLV Sub-Type	TLV Name	TLV Variable	System	LLDP-MED MIB Object
1	LLDP-MED Capabilities	LLDP-MED Capabilities	Local	lldpXMedPortCapSupported
				lldpXMedPortConfigTLVsTx Enable
			Remote	lldpXMedRemCapSupported
				lldpXMedRemConfigTLVsTxEnable
		LLDP-MED Class Type	Local	lldpXMedLocDevice Class
			Remote	lldpXMedRemDevice Class
2	Network Policy	Application Type	Local	lldpXMedLocMediaPolicyAppType
			Remote	lldpXMedRemMediaPolicyAppType
		Unknown Policy Flag	Local	lldpXMedLocMediaPolicyUnknown
			Remote	lldpXMedLocMediaPolicyUnknown
		Tagged Flag	Local	lldpXMedLocMediaPolicyTagged
			Remote	lldpXMedLocMediaPolicyTagged
		VLAN ID	Local	lldpXMedLocMediaPolicyVlanID
			Remote	lldpXMedRemMediaPolicyVlanID
		L2 Priority	Local	lldpXMedLocMediaPolicyPriority
			Remote	lldpXMedRemMediaPolicyPriority
		DSCP Value	Local	lldpXMedLocMediaPolicyDscp
			Remote	lldpXMedRemMediaPolicyDscp

TLV Sub-Type	TLV Name	TLV Variable	System	LLDP-MED MIB Object
3	Location Identifier	Location Data Format	Local	lldpXMedLocLocationSubtype
			Remote	lldpXMedRemLocationSubtype
		Location ID Data	Local	lldpXMedLocLocationInfo
			Remote	lldpXMedRemLocationInfo
4	Extended Power via MDI	Power Device Type	Local	lldpXMedLocXPoEDeviceType
			Remote	lldpXMedRemXPoEDeviceType
		Power Source	Local	lldpXMedLocXPoEPSEPowerSource lldpXMedLocXPoEPDPowerSource
			Remote	lldpXMedRemXPoEPSEPowerSource lldpXMedRemXPoEPDPowerSource
		Power Priority	Local	lldpXMedLocXPoEPDPowerPriority lldpXMedLocXPoEPSEPortPDPriority
			Remote	lldpXMedRemXPoEPSEPowerPriority lldpXMedRemXPoEPDPowerPriority
		Power Value	Local	lldpXMedLocXPoEPSEPortPowerAv lldpXMedLocXPoEPDPowerReq
			Remote	lldpXMedRemXPoEPSEPowerAv lldpXMedRemXPoEPDPowerReq

27

Microsoft Network Load Balancing

This functionality is supported on the S6000 platform.

Network Load Balancing (NLB) is a clustering functionality that is implemented by Microsoft on Windows 2000 Server and Windows Server 2003 operating systems. NLB uses a distributed methodology or pattern to equally split and balance the network traffic load across a set of servers that are part of the cluster or group. NLB combines the servers into a single multicast group and attempts to use the standard multicast IP or unicast IP addresses, and MAC addresses for the transmission of network traffic. At the same time, it also uses a single virtual IP address for all clients as the destination IP address, which enables servers to join the same multicast group in a way that is transparent to the clients (the clients do not notice the addition of new servers to the group). The clients use a cluster IP address to connect to the server. The NLB functionality enables flooding of traffic over the VLAN ports (for unicast mode) or a subset of ports in a VLAN (for multicast mode) to avoid overloading and effective performance of the servers for optimal processing of data packets.

NLB functions in two modes, namely unicast mode and multicast mode. The cluster IP address and the associated cluster MAC address are configured in the NLB application running on the Windows Server. In the unicast mode, when the server IP address is attempted to be resolved to the MAC address using the ARP application, the switch determines whether the ARP reply, obtained from the server, is of an NLB type. The switch then maps the IP address (cluster IP) with the MAC address (cluster MAC address). In multicast mode, the cluster IP address is mapped to a cluster multicast MAC address that is configured using a static ARP CLI configuration command. After the NLB entry is learned, the traffic is forwarded to all the servers in the VLAN corresponding to the cluster virtual IP address.

NLB Unicast Mode Scenario

Consider a sample topology in which four servers, namely S1 through S4, are configured as a cluster or a farm. This set of servers is connected to a Layer 3 switch, which in turn is connected to the end-clients. The servers contain a single IP address (IP-cluster address of 172.16.2.20) and a single unicast MAC address (MAC-Cluster address of 00-bf-ac-10-00-01) for load-balancing. Because multiple ports of a switch cannot learn a single MAC address, the servers are assigned with MAC addresseses of MAC-s1 to MAC-s4) respectively on S1 through S4 in addition to the MAC cluster address. All the servers of the cluster belong to the VLAN named VLAN1.

In unicast NLB mode, the following sequence of events occurs:

- The switch sends an ARP request to resolve the IP address to the cluster MAC address.
- The ARP servers send an ARP response with the MAC cluster address in the ARP header and a MAC address of MAC-s1/s2/s3/s4 (for servers S1 through S4) in the Ethernet header.
- The switch associates the IP address with the MAC cluster address with the last ARP response it obtains. Assume that in this case, the last ARP reply is obtained from MAC-s4.(assuming that the ARP response with MAC-s4 is received as the last one). The interface associated with server, S4, is added to the ARP table.

- With NLB feature enabled, after learning the NLB ARP entry, all the subsequent traffic is flooded on all ports in VLAN1.

With NLB, the data frame is forwarded to all the servers for them to perform load-balancing.

NLB Multicast Mode Scenario

Consider a sample topology in which four servers, namely S1 through S4, are configured as a cluster or a farm. This set of servers is connected to a Layer 3 switch, which in turn is connected to the end-clients. They contain a single multicast MAC address (MAC-Cluster: 03-00-5E-11-11-11).

In the multicast NLB mode, a static ARP configuration command is configured to associate the cluster IP address with a multicast cluster MAC address.

With multicast NLB mode, the data is forwarded to all the servers based on the port specified using the Layer 2 multicast command, which is the `mac-address-table static <multicast_mac> multicast vlan <vlan_id> output-range <port1>, <port2>` command in CONFIGURATION mode.

Limitations With Enabling NLB on Switches

The following limitations apply to switches on which you configure NLB:

- The NLB unicast mode uses switch flooding to transmit all packets to all the servers that are part of the VLAN. When a large volume of traffic is processed, the clustering performance might be impacted in a small way. This limitation is applicable to switches that perform unicast flooding in the software.
- The `ip vlan-flooding` command applies globally across the system and for all VLANs. In cases where the NLB is applicable and the ARP replies contain a discrepancy in the Ethernet SHA and ARP header SHA frames, a flooding of packets over the relevant VLAN occurs.
- The maximum number of concurrent clusters that is supported is eight.

Benefits and Working of Microsoft Clustering

Microsoft clustering allows multiple servers using Microsoft Windows to be represented by one MAC address and IP address in order to provide transparent failover or balancing. Dell Networking OS does not recognize server clusters by default; it must be configured to do so. When an ARP request is sent to a server cluster, either the active server or all the servers send a reply, depending on the cluster configuration. If the active server sends a reply, the Dell switch learns the active server's MAC address. If all servers reply, the switch registers only the last received ARP reply, and the switch learns one server's actual MAC address; the virtual MAC address is never learned. Because the virtual MAC address is never learned, traffic is forwarded to only one server rather than the entire cluster, and failover and balancing are not preserved.

To preserve failover and balancing, the switch forwards the traffic destined for the server cluster to all member ports in the VLAN connected to the cluster. To ensure that this happens, you must configure the `ip vlan-flooding` command on the Dell switch at the time that the Microsoft cluster is configured. The server MAC address is given in the Ethernet frame header of the ARP reply, while the virtual MAC address representing the cluster is given in the payload. Then, all the traffic destined for the cluster is flooded out of all member ports. Since all the servers in the cluster receive traffic, failover and balancing are preserved.

Enable and Disable VLAN Flooding

- The older ARP entries are overwritten whenever newer NLB entries are learned.
- All ARP entries, learned after the feature is enabled, are deleted when the feature is disabled, and RP2 triggers an ARP resolution. The feature is disabled with the `no ip vlan-flooding` command.
- When a port is added to the VLAN, the port automatically receives traffic if the feature is enabled. Old ARP entries are not deleted or updated.
- When a member port is deleted, its ARP entries are also deleted from the CAM.
- Port channels in the VLAN also receive traffic.
- There is no impact on the configuration from saving the configuration.
- The feature, if enabled, is displayed in the `show running-config` command output that displays the `ip vlan-flooding` CLI configuration. Apart from it, there is no indication of the enabling of this capability.

Configuring a Switch for NLB

This functionality is supported on the S6000 platform.

To enable a switch for unicast NLB mode of functioning, perform the following steps:

Enter the `ip vlan-flooding` command to specify that all Layer 3 unicast routed data traffic, going through a VLAN member port, needs to be flooded across all the member ports of that VLAN. There might be some ARP table entries that are resolved through ARP packets, which had the Ethernet MAC SA different from the MAC information inside the ARP packet. This unicast data traffic flooding occurs only for those packets that use these ARP entries.

CONFIGURATION mode

```
ip vlan-flooding
```

To enable a switch for multicast NLB mode of functioning, perform the following steps:

1. In the multicast mode of NLB, add a static ARP entry by entering the `arp ip-address multicast-mac-address` command in Global configuration mode to associate an IP address with a multicast MAC address in the switch. This setting causes the multicast MAC address to be mapped to the cluster IP address for the NLB mode of operation of the switch.

   INTERFACE mode

   ```
   arp ip-address multicast-mac-address interface
   ```
2. Associate specific MAC or hardware addresses to VLANs.

   CONFIGURATION mode

   ```
   mac-address-table static multicast-mac-address vlan vlan-id output-range
   interface
   ```

28

Multicast Source Discovery Protocol (MSDP)

Multicast source discovery protocol (MSDP) is supported on the S6000 platform.

Protocol Overview

MSDP is a Layer 3 protocol that connects IPv4 protocol-independent multicast-sparse mode (PIM-SM) domains. A domain in the context of MSDP is a contiguous set of routers operating PIM within a common boundary defined by an exterior gateway protocol, such as border gateway protocol (BGP).

Each rendezvous point (RP) peers with every other RP via the transmission control protocol (TCP). Through this connection, peers advertise the sources in their domain.

1. When an RP in a PIM-SM domain receives a PIM register message from a source, it sends a source-active (SA) message to MSDP peers, as shown in the following illustration.
2. Each MSDP peer receives and forwards the message to its peers away from the originating RP.
3. When an MSDP peer receives an SA message, it determines if there are any group members within the domain interested in any of the advertised sources. If there are, the receiving RP sends a join message to the originating RP, creating a shortest path tree (SPT) to the source.

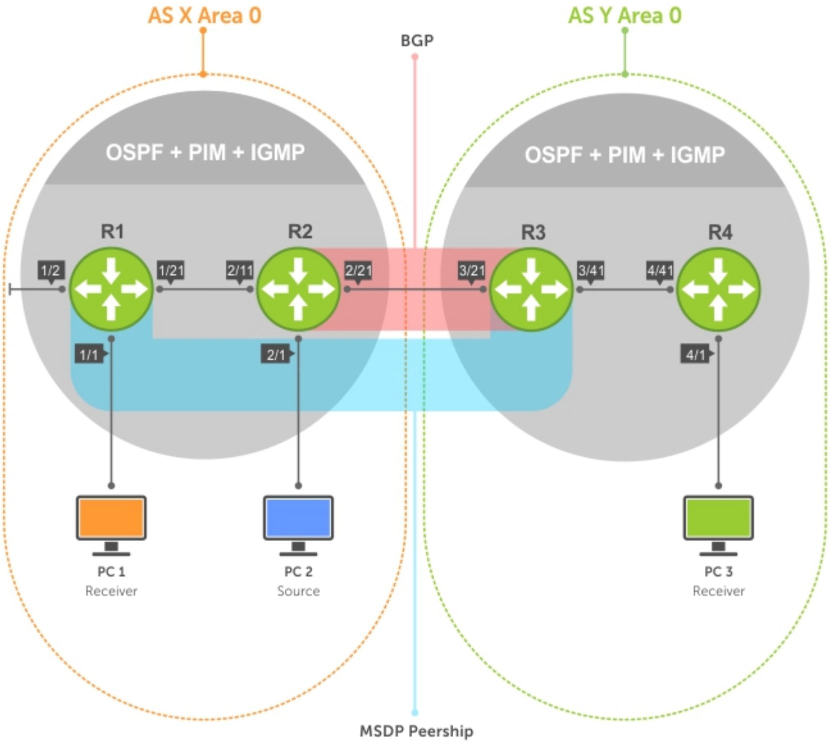

AS X Area 0
BGP
AS Y Area 0
OSPF + PIM + IGMP
OSPF + PIM + IGMP
R1
R2
R3
R4
1/2
1/21
2/11
2/21
3/21
3/41
4/41
1/1
2/1
4/1
PC 1
Receiver
PC 2
Source
PC 3
Receiver
MSDP Peership

Figure 74. Multicast Source Discovery Protocol (MSDP)

RPs advertise each (S,G) in its domain in type, length, value (TLV) format. The total number of TLVs contained in the SA is indicated in the “Entry Count” field. SA messages are transmitted every 60 seconds, and immediately when a new source is detected.

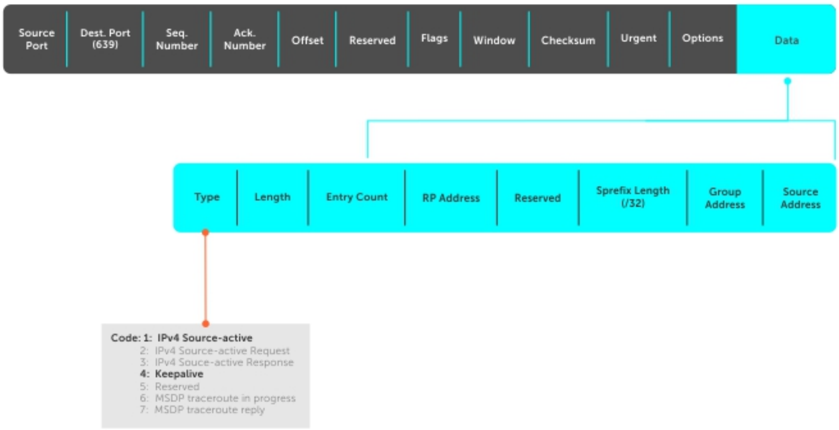

Source Port
Dest. Port (639)
Seq. Number
Ack. Number
Offset
Reserved
Flags
Window
Checksum
Urgent
Options
Data
Type
Length
Entry Count
RP Address
Reserved
Sprefix Length (/32)
Group Address
Source Address
Code: 1: IPv4 Source-active
2: IPv4 Source-active Request
3: IPv4 Souce-active Response
4: Keepalive
5: Reserved
6: MSDP traceroute in progress
7: MSDP traceroute reply

Figure 75. MSDP SA Message Format

Anycast RP

Using MSDP, anycast RP provides load sharing and redundancy in PIM-SM networks. Anycast RP allows two or more rendezvous points (RPs) to share the load for source registration and the ability to act as hot backup routers for each other.

Anycast RP allows you to configure two or more RPs with the same IP address on Loopback interfaces. The Anycast RP Loopback address are configured with a 32-bit mask, making it a host address. All downstream routers are configured to know that the Anycast RP Loopback address is the IP address of their local RP. IP routing automatically selects the closest RP for each source and receiver. Assuming that the sources are evenly spaced around the network, an equal number of sources register with each RP. Consequently, all the RPs in the network share the process of registering the sources equally. Because a source may register with one RP and receivers may join to a different RP, a method is needed for the RPs to exchange information about active sources. This information exchange is done with MSDP.

With Anycast RP, all the RPs are configured to be MSDP peers of each other. When a source registers with one RP, an SA message is sent to the other RPs informing them that there is an active source for a particular multicast group. The result is that each RP is aware of the active sources in the area of the other RPs. If any of the RPs fail, IP routing converges and one of the RPs becomes the active RP in more than one area. New sources register with the backup RP. Receivers join toward the new RP and connectivity is maintained.

Implementation Information

The Dell Networking OS implementation of MSDP is in accordance with RFC 3618 and Anycast RP is in accordance with RFC 3446.

Configure Multicast Source Discovery Protocol

Configuring MSDP is a four-step process.

1. Enable an exterior gateway protocol (EGP) with at least two routing domains.

   Refer to the following figures.

   The MSDP Sample Configurations show the OSPF-BGP configuration used in this chapter for MSDP. Also, refer to Open Shortest Path First (OSPFv2) and Border Gateway Protocol IPv4 (BGPv4).
2. Configure PIM-SM within each EGP routing domain.

   Refer to the following figures.

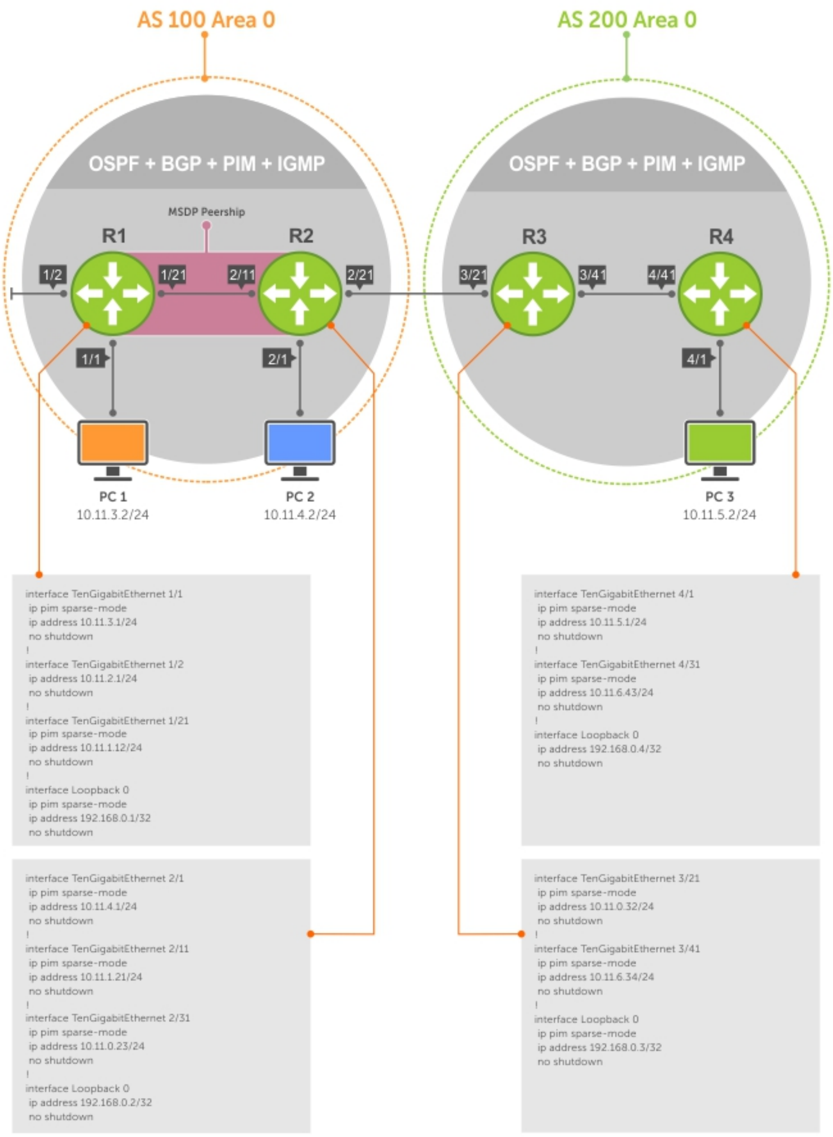
   The MSDP Sample Configurations show the PIM-SM configuration in this chapter for MSDP. Also, refer to PIM Sparse-Mode (PIM-SM).
3. Enable MSDP.
4. Peer the RPs in each routing domain with each other. Refer to Enable MSDP.

Related Configuration Tasks

The following lists related MSDP configuration tasks.

- Enable MSDP
- Manage the Source-Active Cache

- [Accept Source-Active Messages that Fail the RFP Check]
- [Specifying Source-Active Messages]
- [Limiting the Source-Active Cache]
- [Preventing MSDP from Caching a Local Source]
- [Preventing MSDP from Caching a Remote Source]
- [Preventing MSDP from Advertising a Local Source]
- [Terminating a Peership]
- [Clearing Peer Statistics]
- [Debugging MSDP]
- [MSDP with Anycast RP]
- [MSDP Sample Configurations]



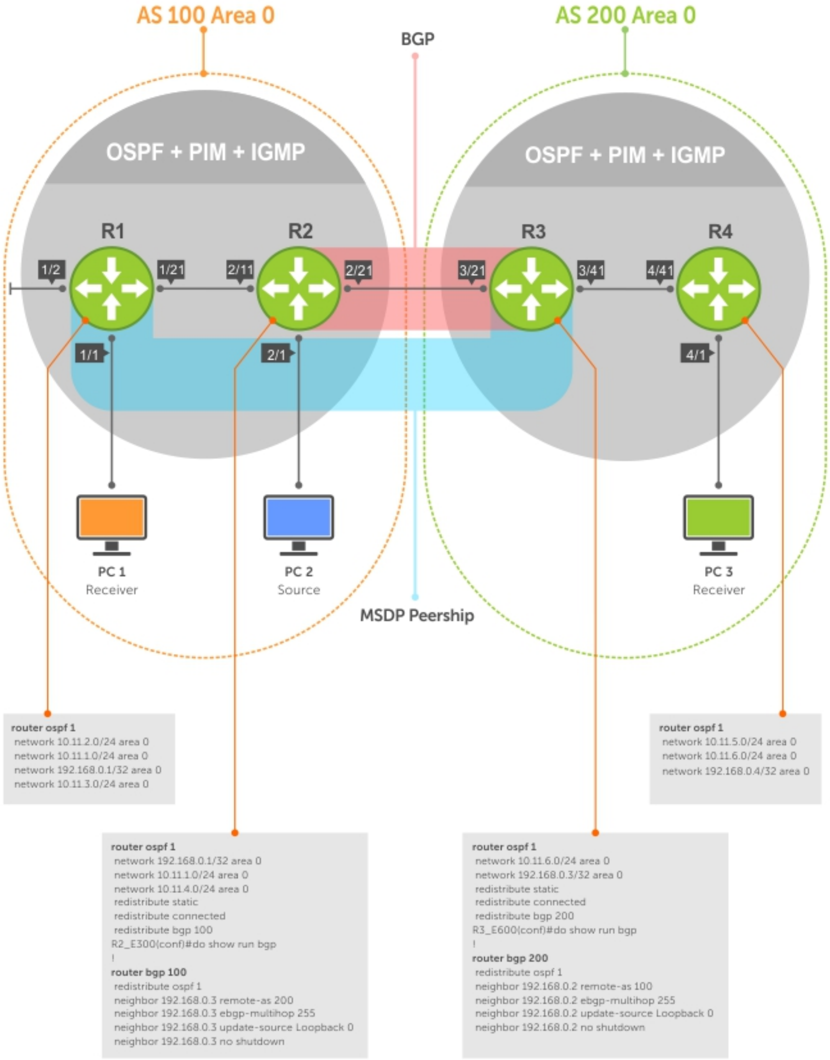
```
interface TenGigabitEthernet 1/1
 ip pim sparse-mode
 ip address 10.11.3.1/24
 no shutdown
!
interface TenGigabitEthernet 1/2
 ip address 10.11.2.1/24
 no shutdown
!
interface TenGigabitEthernet 1/21
 ip pim sparse-mode
 ip address 10.11.1.12/24
 no shutdown
!
interface Loopback 0
 ip pim sparse-mode
 ip address 192.168.0.1/32
 no shutdown
```

```
interface TenGigabitEthernet 2/1
 ip pim sparse-mode
 ip address 10.11.4.1/24
 no shutdown
!
interface TenGigabitEthernet 2/11
 ip pim sparse-mode
 ip address 10.11.1.21/24
 no shutdown
!
interface TenGigabitEthernet 2/31
 ip pim sparse-mode
 ip address 10.11.0.23/24
 no shutdown
!
interface Loopback 0
 ip address 192.168.0.2/32
 no shutdown
```

```
interface TenGigabitEthernet 4/1
 ip pim sparse-mode
 ip address 10.11.5.1/24
 no shutdown
!
interface TenGigabitEthernet 4/31
 ip pim sparse-mode
 ip address 10.11.6.43/24
 no shutdown
!
interface Loopback 0
 ip address 192.168.0.4/32
 no shutdown
```

```
interface TenGigabitEthernet 3/21
 ip pim sparse-mode
 ip address 10.11.0.32/24
 no shutdown
!
interface TenGigabitEthernet 3/41
 ip pim sparse-mode
 ip address 10.11.6.34/24
 no shutdown
!
interface Loopback 0
 ip pim sparse-mode
 ip address 192.168.0.3/32
 no shutdown
```

Figure 76. Configuring Interfaces for MSDP


AS 100 Area 0
BGP
AS 200 Area 0
OSPF + PIM + IGMP
OSPF + PIM + IGMP
R1
R2
R3
R4
1/2
1/21
2/11
2/21
3/21
3/41
4/41
1/1
2/1
4/1
PC 1
Receiver
PC 2
Source
PC 3
Receiver
MSDP Peership

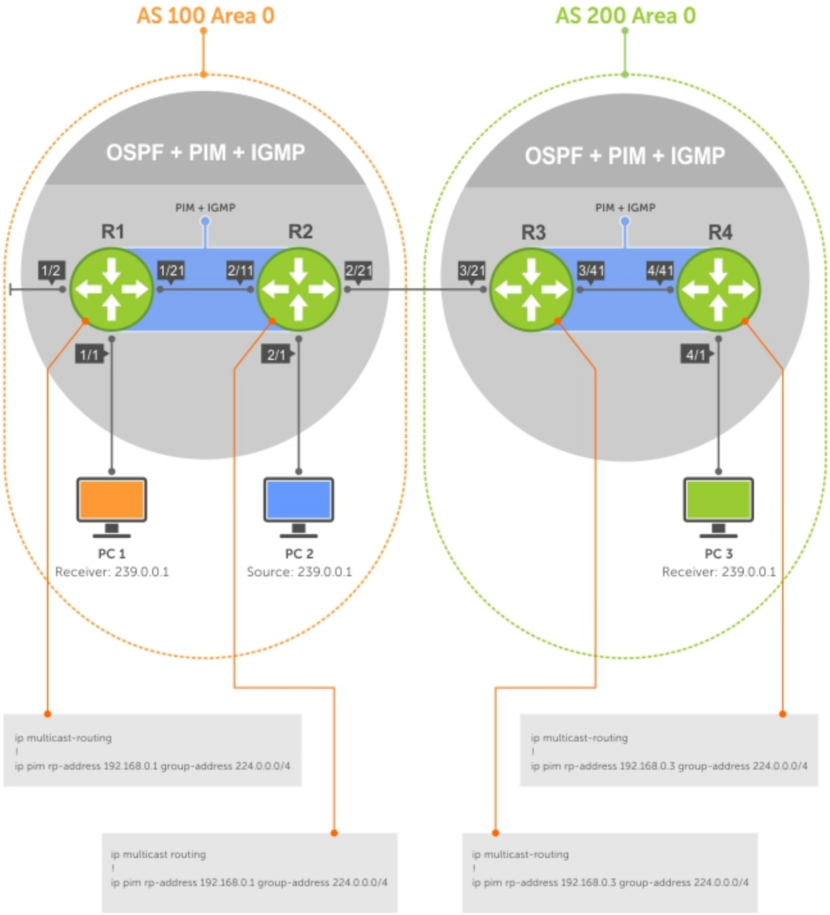
```
router ospf 1
 network 10.11.2.0/24 area 0
 network 10.11.1.0/24 area 0
 network 192.168.0.1/32 area 0
 network 10.11.3.0/24 area 0
```

```
router ospf 1
 network 10.11.5.0/24 area 0
 network 10.11.6.0/24 area 0
 network 192.168.0.4/32 area 0
```

```
router ospf 1
 network 192.168.0.1/32 area 0
 network 10.11.1.0/24 area 0
 network 10.11.4.0/24 area 0
 redistribute static
 redistribute connected
 redistribute bgp 100
R2_E300(conf)#do show run bgp
!
router bgp 100
 redistribute ospf 1
 neighbor 192.168.0.3 remote-as 200
 neighbor 192.168.0.3 ebgp-multihop 255
 neighbor 192.168.0.3 update-source Loopback 0
 neighbor 192.168.0.3 no shutdown
```

```
router ospf 1
 network 10.11.6.0/24 area 0
 network 192.168.0.3/32 area 0
 redistribute static
 redistribute connected
 redistribute bgp 200
R3_E600(conf)#do show run bgp
!
router bgp 200
 redistribute ospf 1
 neighbor 192.168.0.2 remote-as 100
 neighbor 192.168.0.2 ebgp-multihop 255
 neighbor 192.168.0.2 update-source Loopback 0
 neighbor 192.168.0.2 no shutdown
```

Figure 77. Configuring OSPF and BGP for MSDP


AS 100 Area 0
AS 200 Area 0
OSPF + PIM + IGMP
OSPF + PIM + IGMP
PIM + IGMP
PIM + IGMP
R1
R2
R3
R4
1/2
1/21
2/11
2/21
3/21
3/41
4/41
1/1
2/1
4/1
PC 1
Receiver: 239.0.0.1
PC 2
Source: 239.0.0.1
PC 3
Receiver: 239.0.0.1

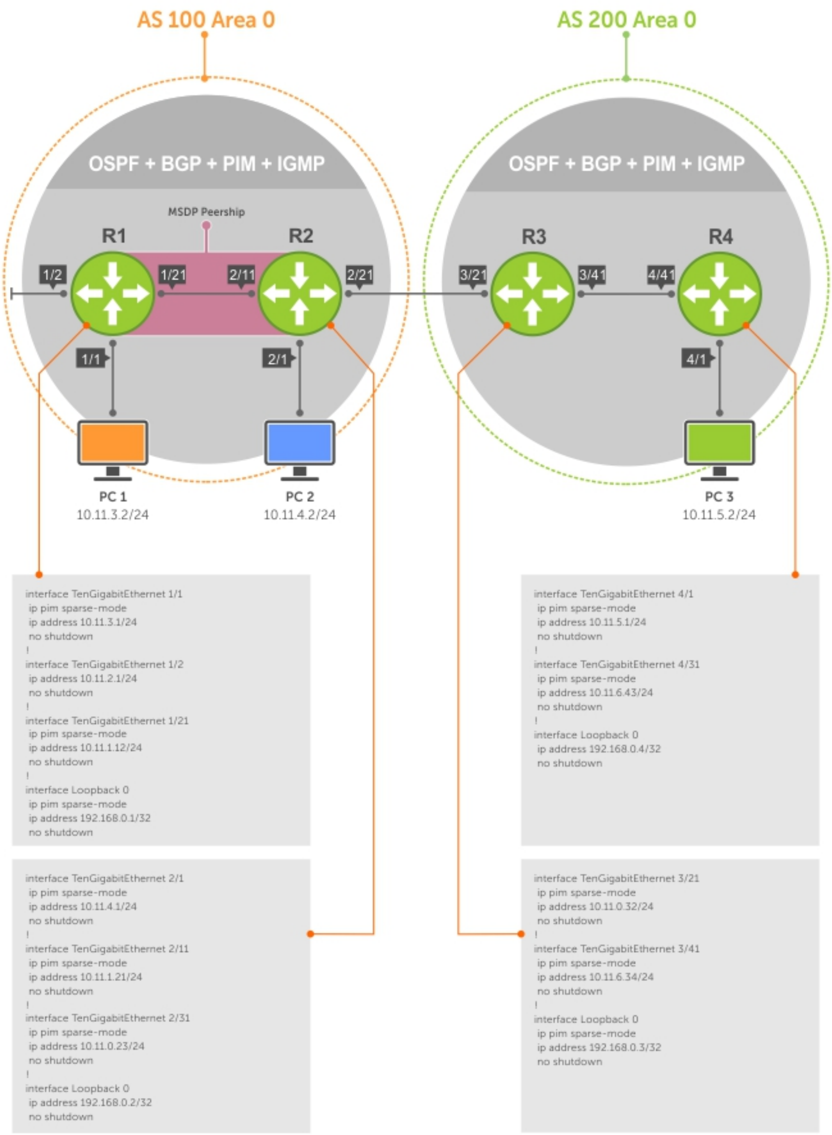
```
ip multicast-routing
!
ip pim rp-address 192.168.0.1 group-address 224.0.0.0/4
```

```
ip multicast-routing
!
ip pim rp-address 192.168.0.3 group-address 224.0.0.0/4
```

```
ip multicast routing
!
ip pim rp-address 192.168.0.1 group-address 224.0.0.0/4
```

```
ip multicast-routing
!
ip pim rp-address 192.168.0.3 group-address 224.0.0.0/4
```

Figure 78. Configuring PIM in Multiple Routing Domains



```
interface TenGigabitEthernet 1/1
 ip pim sparse-mode
 ip address 10.11.3.1/24
 no shutdown
!
interface TenGigabitEthernet 1/2
 ip address 10.11.2.1/24
 no shutdown
!
interface TenGigabitEthernet 1/21
 ip pim sparse-mode
 ip address 10.11.1.12/24
 no shutdown
!
interface Loopback 0
 ip pim sparse-mode
 ip address 192.168.0.1/32
 no shutdown
```

```
interface TenGigabitEthernet 4/1
 ip pim sparse-mode
 ip address 10.11.5.1/24
 no shutdown
!
interface TenGigabitEthernet 4/31
 ip pim sparse-mode
 ip address 10.11.6.43/24
 no shutdown
!
interface Loopback 0
 ip address 192.168.0.4/32
 no shutdown
```

```
interface TenGigabitEthernet 2/1
 ip pim sparse-mode
 ip address 10.11.4.1/24
 no shutdown
!
interface TenGigabitEthernet 2/11
 ip pim sparse-mode
 ip address 10.11.1.21/24
 no shutdown
!
interface TenGigabitEthernet 2/31
 ip pim sparse-mode
 ip address 10.11.0.23/24
 no shutdown
!
interface Loopback 0
 ip address 192.168.0.2/32
 no shutdown
```

```
interface TenGigabitEthernet 3/21
 ip pim sparse-mode
 ip address 10.11.0.32/24
 no shutdown
!
interface TenGigabitEthernet 3/41
 ip pim sparse-mode
 ip address 10.11.6.34/24
 no shutdown
!
interface Loopback 0
 ip pim sparse-mode
 ip address 192.168.0.3/32
 no shutdown
```

Figure 79. Configuring MSDP

Enable MSDP

Enable MSDP by peering RPs in different administrative domains.

1. Enable MSDP.
   CONFIGURATION mode

   ```
   ip multicast-msdp
   ```

2. Peer PIM systems in different administrative domains.
   CONFIGURATION mode

   ```
   ip msdp peer connect-source
   ```

Examples of Configuring and Viewing MSDP

```
    R3_E600(conf)#ip multicast-msdp
    R3_E600(conf)#ip msdp peer 192.168.0.1 connect-source Loopback 0
    R3_E600(conf)#do show ip msdp summary

    Peer Addr   Local Addr   State   Source   SA   Up/Down
Description
```

To view details about a peer, use the `show ip msdp peer` command in EXEC privilege mode.

Multicast sources in remote domains are stored on the RP in the source-active cache (SA cache). The system does not create entries in the multicast routing table until there is a local receiver for the corresponding multicast group.

```
R3_E600#show ip msdp peer

Peer Addr: 192.168.0.1
    Local Addr: 192.168.0.3(639) Connect Source: Lo 0
    State: Established Up/Down Time: 00:15:20
    Timers: KeepAlive 30 sec, Hold time 75 sec
    SourceActive packet count (in/out): 8/0
    SAs learned from this peer: 1
    SA Filtering:
    Input (S,G) filter: none
    Output (S,G) filter: none
```

Manage the Source-Active Cache

Each SA-originating RP caches the sources inside its domain (domain-local), and the sources which it has learned from its peers (domain-remote).

By caching sources:

- domain-local receivers experience a lower join latency
- RPs can transmit SA messages periodically to prevent SA storms
- only sources that are in the cache are advertised in the SA to prevent transmitting multiple copies of the same source information

Viewing the Source-Active Cache

To view the source-active cache, use the following command.

- View the SA cache.

  EXEC Privilege mode

  ```
  show ip msdp sa-cache
  ```

Example of the `show ip msdp sa-cache` Command

```
R3_E600#show ip msdp sa-cache
MSDP Source-Active Cache - 1 entries
GroupAddr  SourceAddr  RPAddr       LearnedFrom  Expire UpTime
239.0.0.1  10.11.4.2   192.168.0.1  192.168.0.1  76     00:10:44
```

Limiting the Source-Active Cache

Set the upper limit of the number of active sources that the Dell Networking OS caches.
The default active source limit is 500K messages. When the total number of active sources reaches the specified limit, subsequent active sources are dropped even if they pass the reverse path forwarding (RPF) and policy check.

To limit the number of sources that SA cache stores, use the following command.

- Limit the number of sources that can be stored in the SA cache.

  EXEC Privilege mode

  ```
  show ip msdp sa-limit
  ```

If the total number of active sources is already larger than the limit when limiting is applied, the sources that are already in Dell Networking OS are not discarded. To enforce the limit in such a situation, use the `clear ip msdp sa-cache` command to clear all existing entries.

Clearing the Source-Active Cache

To clear the source-active cache, use the following command.

- Clear the SA cache of all, local, or rejected entries, or entries for a specific group.

  CONFIGURATION mode

  ```
  clear ip msdp sa-cache [group-address | local | rejected-sa]
  ```

Enabling the Rejected Source-Active Cache

To cache rejected sources, use the following command.
Active sources can be rejected because the RPF check failed, the SA limit is reached, the peer RP is unreachable, or the SA message has a format error.

- Cache rejected sources.

  CONFIGURATION mode

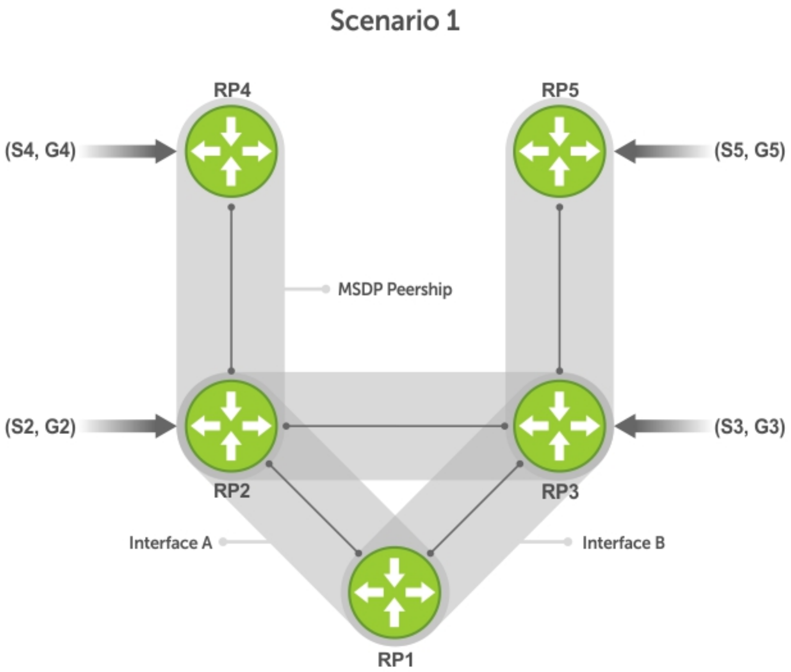
  ```
  ip msdp cache-rejected-sa
  ```

Accept Source-Active Messages that Fail the RFP Check

A default peer is a peer from which active sources are accepted even though they fail the RFP check.

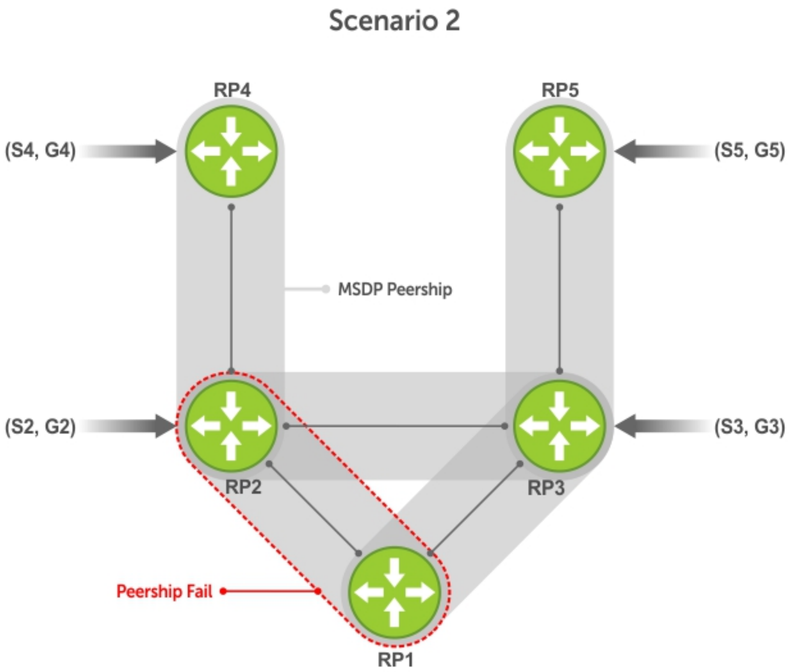
Referring to the following illustrations:

- In Scenario 1, all MSPD peers are up.
- In Scenario 2, the peership between RP1 and RP2 is down, but the link (and routing protocols) between them is still up. In this case, RP1 learns all active sources from RP3, but the sources from RP2 and RP4 are rejected because the reverse path to these routers is through Interface A.
- In Scenario 3, RP3 is configured as a default MSDP peer for RP1 and so the RPF check is disregarded for RP3.
- In Scenario 4, RP1 has a default peer plus an access list. The list permits RP4 so the RPF check is disregarded for active sources from it, but RP5 (and all others because of the implicit deny all) are subject to the RPF check and fail, so those active sources are rejected.


Scenario 1
RP4
RP5
(S4, G4)
(S5, G5)
MSDP Peership
(S2, G2)
(S3, G3)
RP2
RP3
Interface A
Interface B
RP1

Group	Source	RP	Peer
G2	S2	RP2	R2
G3	S3	RP3	R3
G4	S4	RP4	R4
G5	S5	RP5	R5

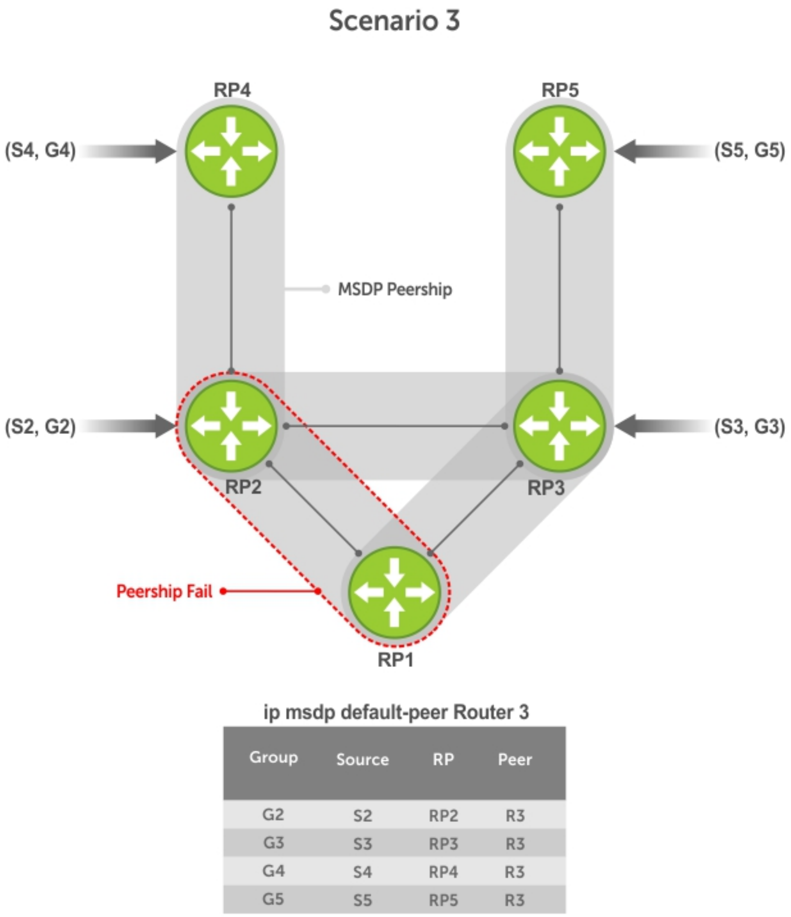
Figure 80. MSDP Default Peer, Scenario 1


Scenario 2
RP4
RP5
(S4, G4)
(S5, G5)
MSDP Peership
(S2, G2)
(S3, G3)
RP2
RP3
Peership Fail
RP1

ip msdp default-peer Router 3

Group	Source	RP	Peer
G2	S2	RP2	R3 RPF-Fail
G3	S3	RP3	R3
G4	S4	RP4	R3 RPF-Fail
G5	S5	RP5	R3

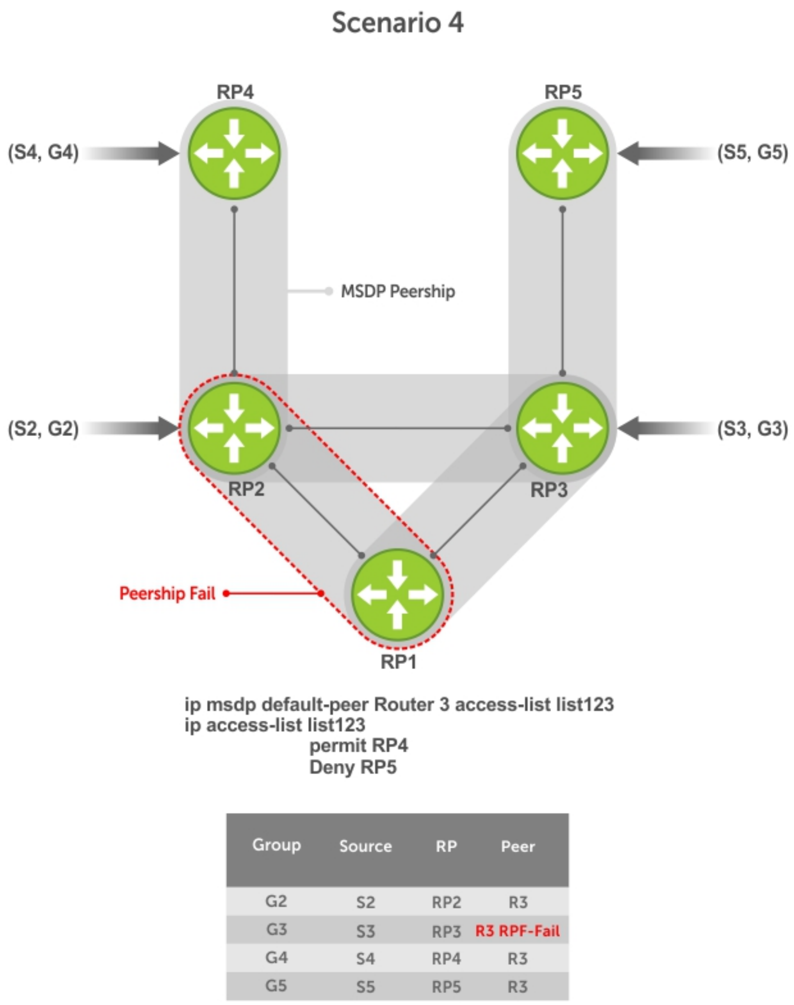
Figure 81. MSDP Default Peer, Scenario 2


Scenario 3
RP4
RP5
(S4, G4)
(S5, G5)
MSDP Peership
(S2, G2)
(S3, G3)
RP2
RP3
Peership Fail
RP1

ip msdp default-peer Router 3

Group	Source	RP	Peer
G2	S2	RP2	R3
G3	S3	RP3	R3
G4	S4	RP4	R3
G5	S5	RP5	R3

Figure 82. MSDP Default Peer, Scenario 3



```
ip msdp default-peer Router 3 access-list list123
ip access-list list123
                 permit RP4
                 Deny RP5
```

Group	Source	RP	Peer
G2	S2	RP2	R3
G3	S3	RP3	R3 RPF-Fail
G4	S4	RP4	R3
G5	S5	RP5	R3

Figure 83. MSDP Default Peer, Scenario 4

Specifying Source-Active Messages

To specify messages, use the following command.

- Specify the forwarding-peer and originating-RP from which all active sources are accepted without regard for the RPF check.

  CONFIGURATION mode

  `ip msdp default-peer` *`ip-address`* `list`

  If you do not specify an access list, the peer accepts all sources that peer advertises. All sources from RPs that the ACL denies are subject to the normal RPF check.

Example of the `ip msdp default-peer` Command and Viewing Denied Sources

```
Dell(conf)#ip msdp peer 10.0.50.2 connect-source Vlan 50
Dell(conf)#ip msdp default-peer 10.0.50.2 list fifty
Dell(conf)#ip access-list standard fifty
Dell(conf)#seq 5 permit host 200.0.0.50

Dell#ip msdp sa-cache
MSDP Source-Active Cache - 3 entries
GroupAddr   SourceAddr  RPAddr      LearnedFrom  Expire  UpTime
229.0.50.2  24.0.50.2   200.0.0.50  10.0.50.2    73      00:13:49
229.0.50.3  24.0.50.3   200.0.0.50  10.0.50.2    73      00:13:49
229.0.50.4  24.0.50.4   200.0.0.50  10.0.50.2    73      00:13:49

Dell#ip msdp sa-cache rejected-sa
MSDP Rejected SA Cache
  3 rejected SAs received, cache-size 32766
UpTime    GroupAddr    SourceAddr  RPAddr      LearnedFrom  Reason
00:33:18  229.0.50.64  24.0.50.64  200.0.1.50  10.0.50.2    Rpf-Fail
00:33:18  229.0.50.65  24.0.50.65  200.0.1.50  10.0.50.2    Rpf-Fail
00:33:18  229.0.50.66  24.0.50.66  200.0.1.50  10.0.50.2    Rpf-Fail
```

Limiting the Source-Active Messages from a Peer

To limit the source-active messages from a peer, use the following commands.

1. OPTIONAL: Store sources that are received after the limit is reached in the rejected SA cache.
   CONFIGURATION mode

   `ip msdp cache-rejected-sa`
2. Set the upper limit for the number of sources allowed from an MSDP peer.
   CONFIGURATION mode

   `ip msdp peer` *`peer-address`* `sa-limit`

   The default limit is **100K**.

If the total number of sources received from the peer is already larger than the limit when this configuration is applied, those sources are not discarded. To enforce the limit in such a situation, first clear the SA cache.

Preventing MSDP from Caching a Local Source

You can prevent MSDP from caching an active source based on source and/or group. Because the source is not cached, it is not advertised to remote RPs.

1. OPTIONAL: Cache sources that are denied by the redistribute list in the rejected SA cache.
   CONFIGURATION mode

   `ip msdp cache-rejected-sa`
2. Prevent the system from caching local SA entries based on source and group using an extended ACL.
   CONFIGURATION mode

   `ip msdp redistribute list`

Example of Verifying the System is not Caching Local Sources

When you apply this filter, the SA cache is not affected immediately. When sources that are denied by the ACL time out, they are not refreshed. Until they time out, they continue to reside in the cache. To apply the redistribute filter to entries already present in the SA cache, first clear the SA cache. You may optionally store denied sources in the rejected SA cache.

```
R1_E600(conf)#do show run msdp
!
ip multicast-msdp
ip msdp peer 192.168.0.3 connect-source Loopback 0
ip msdp redistribute list mylocalfilter
ip msdp cache-rejected-sa 1000
R1_E600(conf)#do show run acl
!
ip access-list extended mylocalfilter
  seq 5 deny ip host 239.0.0.1 host 10.11.4.2
  seq 10 deny ip any any
R1_E600(conf)#do show ip msdp sa-cache
R1_E600(conf)#do show ip msdp sa-cache rejected-sa
MSDP Rejected SA Cache
  1 rejected SAs received, cache-size 1000
UpTime    GroupAddr  SourceAddr  RPAddr       LearnedFrom  Reason
00:02:20  239.0.0.1  10.11.4.2   192.168.0.1  local        Redistribute
```

Preventing MSDP from Caching a Remote Source

To prevent MSDP from caching a remote source, use the following commands.

1. OPTIONAL: Cache sources that the SA filter denies in the rejected SA cache.

   CONFIGURATION mode

   ```
   ip msdp cache-rejected-sa
   ```
2. Prevent the system from caching remote sources learned from a specific peer based on source and group.

   CONFIGURATION mode

   ```
   ip msdp sa-filter list out peer list ext-acl
   ```

Example of Verifying the System is not Caching Remote Sources

As shown in the following example, R1 is advertising source 10.11.4.2. It is already in the SA cache of R3 when an ingress SA filter is applied to R3. The entry remains in the SA cache until it expires and is *not* stored in the rejected SA cache.

```
[Router 3]
R3_E600(conf)#do show run msdp
!
ip multicast-msdp
ip msdp peer 192.168.0.1 connect-source Loopback 0
ip msdp sa-filter in 192.168.0.1 list myremotefilter
R3_E600(conf)#do show run acl
!
ip access-list extended myremotefilter
  seq 5 deny ip host 239.0.0.1 host 10.11.4.2
R3_E600(conf)#do show ip msdp sa-cache
MSDP Source-Active Cache - 1 entries
GroupAddr  SourceAddr  RPAddr       LearnedFrom  Expire  UpTime
239.0.0.1  10.11.4.2   192.168.0.1  192.168.0.1  1       00:03:59
```

```
R3_E600(conf)#do show ip msdp sa-cache
R3_E600(conf)#
R3_E600(conf)#do show ip msdp peer

Peer Addr: 192.168.0.1
    Local Addr: 0.0.0.0(639) Connect Source: Lo 0
    State: Listening Up/Down Time: 00:01:19
    Timers: KeepAlive 30 sec, Hold time 75 sec
    SourceActive packet count (in/out): 0/0
    SAs learned from this peer: 0
    SA Filtering:
    Input (S,G) filter: myremotefilter
    Output (S,G) filter: none
```

Preventing MSDP from Advertising a Local Source

To prevent MSDP from advertising a local source, use the following command.

- Prevent an RP from advertising a source in the SA cache.
  CONFIGURATION mode

  `ip msdp sa-filter list in` *`peer`* `list` *`ext-acl`*

Example of Verifying the System is not Advertising Local Sources

In the following example, R1 stops advertising source 10.11.4.2. Because it is already in the SA cache of R3, the entry remains there until it expires.

```
[Router 1]
R1_E600(conf)#do show run msdp
!
ip multicast-msdp
ip msdp peer 192.168.0.3 connect-source Loopback 0
ip msdp sa-filter out 192.168.0.3 list mylocalfilter
R1_E600(conf)#do show run acl
!
ip access-list extended mylocalfilter
  seq 5 deny ip host 239.0.0.1 host 10.11.4.2
  seq 10 deny ip any any
R1_E600(conf)#do show ip msdp sa-cache
MSDP Source-Active Cache - 1 entries
GroupAddr  SourceAddr  RPAddr       LearnedFrom  Expire  UpTime
239.0.0.1  10.11.4.2   192.168.0.1  local        70      00:27:20
R3_E600(conf)#do show ip msdp sa-cache
MSDP Source-Active Cache - 1 entries
GroupAddr  SourceAddr  RPAddr       LearnedFrom  Expire  UpTime
239.0.0.1  10.11.4.2   192.168.0.1  192.168.0.1  1       00:10:29

[Router 3]
R3_E600(conf)#do show ip msdp sa-cache
R3_E600(conf)#
```

To display the configured SA filters for a peer, use the `show ip msdp peer` command from EXEC Privilege mode.

Logging Changes in Peership States

To log changes in peership states, use the following command.

- Log peership state changes.
  CONFIGURATION mode

```
ip msdp log-adjacency-changes
```

Terminating a Peership

MSDP uses TCP as its transport protocol. In a peering relationship, the peer with the lower IP address initiates the TCP session, while the peer with the higher IP address listens on port 639.

- Terminate the TCP connection with a peer.
  CONFIGURATION mode

```
ip msdp shutdown
```

Example of the Verifying that Peering State is Disabled

After the relationship is terminated, the peering state of the terminator is SHUTDOWN, while the peering state of the peer is INACTIVE.

```
[Router 3]
R3_E600(conf)#ip msdp shutdown 192.168.0.1
R3_E600(conf)#do show ip msdp peer

Peer Addr: 192.168.0.1
    Local Addr: 0.0.0.0(0) Connect Source: Lo 0
    State: Shutdown Up/Down Time: 00:00:18
    Timers: KeepAlive 30 sec, Hold time 75 sec
    SourceActive packet count (in/out): 0/0
    SAs learned from this peer: 0
   SA Filtering:
   Input (S,G) filter: myremotefilter
   Output (S,G) filter: none
[Router 1]
R1_E600(conf)#do show ip msdp peer

Peer Addr: 192.168.0.3
    Local Addr: 0.0.0.0(0) Connect Source: Lo 0
    State: Inactive Up/Down Time: 00:00:03
    Timers: KeepAlive 30 sec, Hold time 75 sec
    SourceActive packet count (in/out): 0/0
    SAs learned from this peer: 0
    SA Filtering:
```

Clearing Peer Statistics

To clear the peer statistics, use the following command.

- Reset the TCP connection to the peer and clear all peer statistics.
  CONFIGURATION mode

```
clear ip msdp peer peer-address
```

Example of the `clear ip msdp peer` Command and Verifying Statistics are Cleared

```
R3_E600(conf)#do show ip msdp peer

Peer Addr: 192.168.0.1
    Local Addr: 192.168.0.3(639) Connect Source: Lo 0
    State: Established Up/Down Time: 00:04:26
    Timers: KeepAlive 30 sec, Hold time 75 sec
    SourceActive packet count (in/out): 5/0
    SAs learned from this peer: 0
    SA Filtering:
    Input (S,G) filter: myremotefilter
    Output (S,G) filter: none
R3_E600(conf)#do clear ip msdp peer 192.168.0.1
R3_E600(conf)#do show ip msdp peer

Peer Addr: 192.168.0.1
    Local Addr: 0.0.0.0(0) Connect Source: Lo 0
    State: Inactive Up/Down Time: 00:00:04
    Timers: KeepAlive 30 sec, Hold time 75 sec
    SourceActive packet count (in/out): 0/0
    SAs learned from this peer: 0
    SA Filtering:
    Input (S,G) filter: myremotefilter
    Output (S,G) filter: none
```

Debugging MSDP

To debug MSDP, use the following command.

- Display the information exchanged between peers.

  CONFIGURATION mode

  ```
  debug ip msdp
  ```

Example of the `debug ip msdp` Command

```
R1_E600(conf)#do debug ip msdp
All MSDP debugging has been turned on
R1_E600(conf)#03:16:08 : MSDP-0: Peer 192.168.0.3, sent Keepalive msg
03:16:09 : MSDP-0: Peer 192.168.0.3,  rcvd Keepalive msg
03:16:27 : MSDP-0: Peer 192.168.0.3,  sent Source Active msg
03:16:38 : MSDP-0: Peer 192.168.0.3,  sent Keepalive msg
03:16:39 : MSDP-0: Peer 192.168.0.3,  rcvd Keepalive msg
03:17:09 : MSDP-0: Peer 192.168.0.3,  sent Keepalive msg
03:17:10 : MSDP-0: Peer 192.168.0.3,  rcvd Keepalive msg
03:17:27 : MSDP-0: Peer 192.168.0.3,  sent Source Active msg
Input (S,G) filter: none
  Output (S,G) filter: none
```

MSDP with Anycast RP

Anycast RP uses MSDP with PIM-SM to allow more than one active group to use RP mapping.

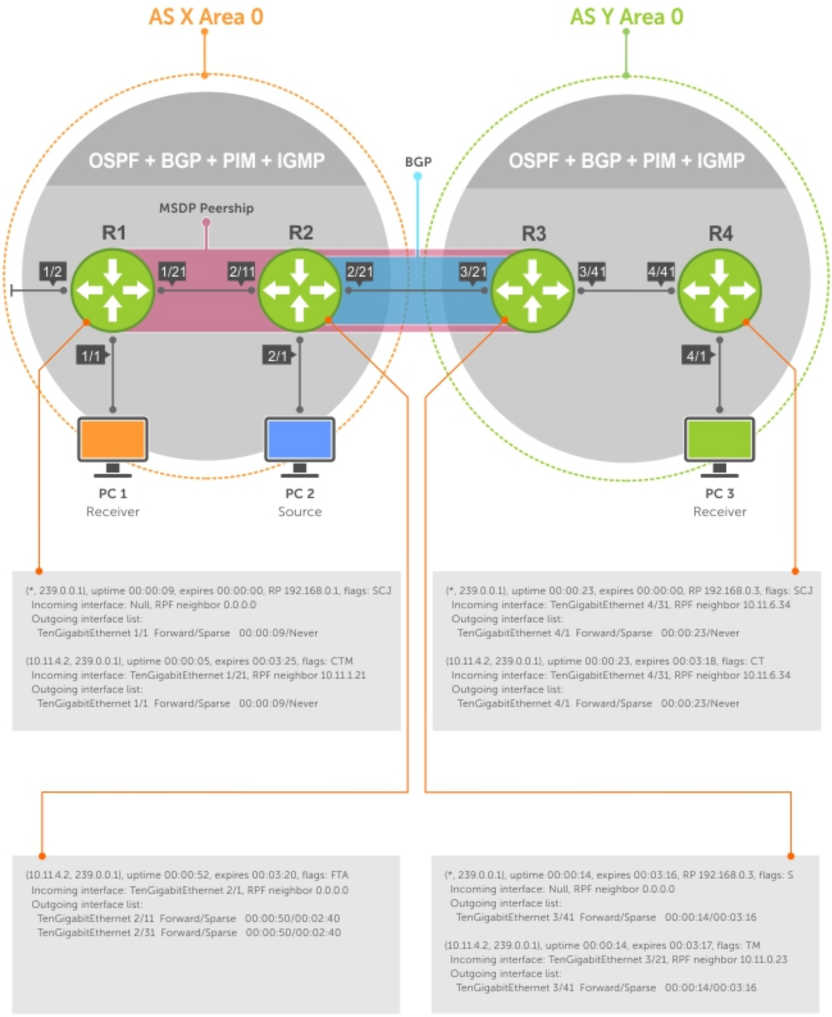
PIM-SM allows only active groups to use RP mapping, which has several implications:

- **traffic concentration**: PIM-SM allows only one active group to RP mapping which means that all traffic for the group must, at least initially, travel over the same part of the network. You can load balance source registration between multiple RPs by strategically mapping groups to RPs, but this technique is less effective as traffic increases because preemptive load balancing requires prior knowledge of traffic distributions.

- **lack of scalable register decasulation**: With only a single RP per group, all joins are sent to that RP regardless of the topological distance between the RP, sources, and receivers, and data is transmitted to the RP until the SPT switch threshold is reached.
- **slow convergence when an active RP fails**: When you configure multiple RPs, there can be considerable convergence delay involved in switching to the backup RP.

Anycast RP relieves these limitations by allowing multiple RPs per group, which can be distributed in a topologically significant manner according to the locations of the sources and receivers.

1. All the RPs serving a given group are configured with an identical anycast address.
2. Sources then register with the topologically closest RP.
3. RPs use MSDP to peer with each other using a unique address.



```
(*, 239.0.0.1), uptime 00:00:09, expires 00:00:00, RP 192.168.0.1, flags: SCJ
  Incoming interface: Null, RPF neighbor 0.0.0.0
  Outgoing interface list:
    TenGigabitEthernet 1/1  Forward/Sparse   00:00:09/Never

(10.11.4.2, 239.0.0.1), uptime 00:00:05, expires 00:03:25, flags: CTM
  Incoming interface: TenGigabitEthernet 1/21, RPF neighbor 10.11.1.21
  Outgoing interface list:
    TenGigabitEthernet 1/1  Forward/Sparse   00:00:09/Never
```

```
(*, 239.0.0.1), uptime 00:00:23, expires 00:00:00, RP 192.168.0.3, flags: SCJ
  Incoming interface: TenGigabitEthernet 4/31, RPF neighbor 10.11.6.34
  Outgoing interface list:
    TenGigabitEthernet 4/1  Forward/Sparse   00:00:23/Never

(10.11.4.2, 239.0.0.1), uptime 00:00:23, expires 00:03:18, flags: CT
  Incoming interface: TenGigabitEthernet 4/31, RPF neighbor 10.11.6.34
  Outgoing interface list:
    TenGigabitEthernet 4/1  Forward/Sparse   00:00:23/Never
```

```
(10.11.4.2, 239.0.0.1), uptime 00:00:52, expires 00:03:20, flags: FTA
  Incoming interface: TenGigabitEthernet 2/1, RPF neighbor 0.0.0.0
  Outgoing interface list:
    TenGigabitEthernet 2/11  Forward/Sparse   00:00:50/00:02:40
    TenGigabitEthernet 2/31  Forward/Sparse   00:00:50/00:02:40
```

```
(*, 239.0.0.1), uptime 00:00:14, expires 00:03:16, RP 192.168.0.3, flags: S
  Incoming interface: Null, RPF neighbor 0.0.0.0
  Outgoing interface list:
    TenGigabitEthernet 3/41  Forward/Sparse   00:00:14/00:03:16

(10.11.4.2, 239.0.0.1), uptime 00:00:14, expires 00:03:17, flags: TM
  Incoming interface: TenGigabitEthernet 3/21, RPF neighbor 10.11.0.23
  Outgoing interface list:
    TenGigabitEthernet 3/41  Forward/Sparse   00:00:14/00:03:16
```

Figure 84. MSDP with Anycast RP

Configuring Anycast RP

To configure anycast RP, use the following commands.

1. In each routing domain that has multiple RPs serving a group, create a Loopback interface on each RP serving the group with the same IP address.

   CONFIGURATION mode

   ```
   interface loopback
   ```

2. Make this address the RP for the group.

   CONFIGURATION mode

   ```
   ip pim rp-address
   ```

3. In each routing domain that has multiple RPs serving a group, create another Loopback interface on each RP serving the group with a unique IP address.

   CONFIGURATION mode

   ```
   interface loopback
   ```

4. Peer each RP with every other RP using MSDP, specifying the unique Loopback address as the connect-source.

   CONFIGURATION mode

   ```
   ip msdp peer
   ```

5. Advertise the network of each of the unique Loopback addresses throughout the network.

   ROUTER OSPF mode

   ```
   network
   ```

Reducing Source-Active Message Flooding

RPs flood source-active messages to all of their peers away from the RP.
When multiple RPs exist within a domain, the RPs forward received active source information back to the originating RP, which violates the RFP rule. You can prevent this unnecessary flooding by creating a mesh-group. A mesh in this context is a topology in which each RP in a set of RPs has a peership with all other RPs in the set. When an RP is a member of the mesh group, it forwards active source information only to its peers outside of the group.

To create a mesh group, use the following command.

- Create a mesh group.

  CONFIGURATION mode

  ```
  ip msdp mesh-group
  ```

Specifying the RP Address Used in SA Messages

The default originator-id is the address of the RP that created the message. In the case of Anycast RP, there are multiple RPs all with the same address.
To use the (unique) address of another interface as the originator-id, use the following command.

- Use the address of another interface as the originator-id instead of the RP address.

CONFIGURATION mode

```
ip msdp originator-id
```

Examples of R1, R2, and R3 Configuration for MSDP with Anycast RP

The following example shows an R1 configuration for MSDP with Anycast RP.

```
ip multicast-routing
!
interface GigabitEthernet 1/1
  ip pim sparse-mode
  ip address 10.11.3.1/24
  no shutdown
!
interface GigabitEthernet 1/2
  ip address 10.11.2.1/24
  no shutdown
!
interface GigabitEthernet 1/21
  ip pim sparse-mode
  ip address 10.11.1.12/24
  no shutdown
!
interface Loopback 0
  ip pim sparse-mode
  ip address 192.168.0.1/32
  no shutdown
!
interface Loopback 1
  ip address 192.168.0.11/32
  no shutdown
!
router ospf 1
  network 10.11.2.0/24 area 0
  network 10.11.1.0/24 area 0
  network 10.11.3.0/24 area 0
  network 192.168.0.11/32 area 0
!
ip multicast-msdp
ip msdp peer 192.168.0.3 connect-source Loopback 1
ip msdp peer 192.168.0.22 connect-source Loopback 1
ip msdp mesh-group AS100 192.168.0.22
ip msdp originator-id Loopback 1!

ip pim rp-address 192.168.0.1 group-address 224.0.0.0/4
```

The following example shows an R2 configuration for MSDP with Anycast RP.

```
ip multicast-routing
!
interface GigabitEthernet 2/1
  ip pim sparse-mode
  ip address 10.11.4.1/24
  no shutdown
!
interface GigabitEthernet 2/11
  ip pim sparse-mode
  ip address 10.11.1.21/24
  no shutdown
!
interface GigabitEthernet 2/31
  ip pim sparse-mode
  ip address 10.11.0.23/24
  no shutdown
!
interface Loopback 0
  ip pim sparse-mode
  ip address 192.168.0.1/32
  no shutdown
!
interface Loopback 1
  ip address 192.168.0.22/32
  no shutdown
!
router ospf 1
  network 10.11.1.0/24 area 0
  network 10.11.4.0/24 area 0
  network 192.168.0.22/32 area 0
  redistribute static
  redistribute connected
  redistribute bgp 100
!
router bgp 100
  redistribute ospf 1
  neighbor 192.168.0.3 remote-as 200
  neighbor 192.168.0.3 ebgp-multihop 255
  neighbor 192.168.0.3 no shutdown
!
ip multicast-msdp
ip msdp peer 192.168.0.3 connect-source Loopback 1
ip msdp peer 192.168.0.11 connect-source Loopback 1
ip msdp mesh-group AS100 192.168.0.11
ip msdp originator-id Loopback 1
!
ip route 192.168.0.3/32 10.11.0.32
!
ip pim rp-address 192.168.0.1 group-address 224.0.0.0/4
```

The following example shows an R3 configuration for MSDP with Anycast RP.

```
ip multicast-routing
!
interface GigabitEthernet 3/21
  ip pim sparse-mode
  ip address 10.11.0.32/24
  no shutdown

interface GigabitEthernet 3/41
  ip pim sparse-mode
  ip address 10.11.6.34/24
  no shutdown
!
interface Loopback 0
  ip pim sparse-mode
  ip address 192.168.0.3/32
  no shutdown
!
router ospf 1
  network 10.11.6.0/24 area 0
  network 192.168.0.3/32 area 0
  redistribute static
  redistribute connected
  redistribute bgp 200
!
router bgp 200
  redistribute ospf 1
  neighbor 192.168.0.22 remote-as 100
  neighbor 192.168.0.22 ebgp-multihop 255
  neighbor 192.168.0.22 update-source Loopback 0
  neighbor 192.168.0.22 no shutdown
!
ip multicast-msdp
ip msdp peer 192.168.0.11 connect-source Loopback 0
ip msdp peer 192.168.0.22 connect-source Loopback 0
ip msdp sa-filter out 192.168.0.22
!
ip route 192.168.0.1/32 10.11.0.23
ip route 192.168.0.22/32 10.11.0.23
!
ip pim rp-address 192.168.0.3 group-address 224.0.0.0/4
```

MSDP Sample Configurations

The following examples show the running-configurations described in this chapter.

For more information, refer to the illustrations in the Related Configuration Tasks section.

MSDP Sample Configuration: R1 Running-Config

MSDP Sample Configuration: R2 Running-Config

MSDP Sample Configuration: R3 Running-Config

MSDP Sample Configuration: R4 Running-Config

```
ip multicast-routing
!
interface GigabitEthernet 1/1
  ip pim sparse-mode
  ip address 10.11.3.1/24
  no shutdown
!
interface GigabitEthernet 1/2
  ip address 10.11.2.1/24
  no shutdown
!
interface GigabitEthernet 1/21
  ip pim sparse-mode
  ip address 10.11.1.12/24
  no shutdown
!
interface Loopback 0
  ip pim sparse-mode
  ip address 192.168.0.1/32
  no shutdown
!
router ospf 1
  network 10.11.2.0/24 area 0
  network 10.11.1.0/24 area 0
  network 192.168.0.1/32 area 0
  network 10.11.3.0/24 area 0
!
ip multicast-msdp
ip msdp peer 192.168.0.3 connect-source Loopback 0
!
ip pim rp-address 192.168.0.1 group-address 224.0.0.0/4
```

```
ip multicast-routing
!
```

```
interface GigabitEthernet 2/1
  ip pim sparse-mode
  ip address 10.11.4.1/24
  no shutdown
!
interface GigabitEthernet 2/11
  ip pim sparse-mode
  ip address 10.11.1.21/24
  no shutdown
!
interface GigabitEthernet 2/31
  ip pim sparse-mode
  ip address 10.11.0.23/24
  no shutdown
!
interface Loopback 0
  ip address 192.168.0.2/32
  no shutdown
!
router ospf 1
  network 10.11.1.0/24 area 0
  network 10.11.4.0/24 area 0
  network 192.168.0.2/32 area 0
  redistribute static
  redistribute connected
  redistribute bgp 100
!
router bgp 100
  redistribute ospf 1
  neighbor 192.168.0.3 remote-as 200
  neighbor 192.168.0.3 ebgp-multihop 255
  neighbor 192.168.0.3 update-source Loopback 0
  neighbor 192.168.0.3 no shutdown
!
ip route 192.168.0.3/32 10.11.0.32
!
ip pim rp-address 192.168.0.1 group-address 224.0.0.0/4
```

```
ip multicast-routing
!
interface GigabitEthernet 3/21
  ip pim sparse-mode
  ip address 10.11.0.32/24
  no shutdown
!
interface GigabitEthernet 3/41
  ip pim sparse-mode
  ip address 10.11.6.34/24
  no shutdown
!
interface ManagementEthernet 0/0
  ip address 10.11.80.3/24
  no shutdown
!
interface Loopback 0
  ip pim sparse-mode
  ip address 192.168.0.3/32
  no shutdown
!
router ospf 1
  network 10.11.6.0/24 area 0
  network 192.168.0.3/32 area 0
  redistribute static
  redistribute connected
  redistribute bgp 200
!
router bgp 200
  redistribute ospf 1
  neighbor 192.168.0.2 remote-as 100
  neighbor 192.168.0.2 ebgp-multihop 255
  neighbor 192.168.0.2 update-source Loopback 0
  neighbor 192.168.0.2 no shutdown
!
ip multicast-msdp
ip msdp peer 192.168.0.1 connect-source Loopback 0
!
ip route 192.168.0.2/32 10.11.0.23


ip multicast-routing
!
interface GigabitEthernet 4/1
  ip pim sparse-mode
  ip address 10.11.5.1/24
  no shutdown
!
interface GigabitEthernet 4/22
  ip address 10.10.42.1/24
  no shutdown
!
interface GigabitEthernet 4/31
  ip pim sparse-mode
  ip address 10.11.6.43/24
  no shutdown
!
interface Loopback 0
  ip address 192.168.0.4/32
  no shutdown
!
router ospf 1
  network 10.11.5.0/24 area 0
  network 10.11.6.0/24 area 0
  network 192.168.0.4/32 area 0
!
ip pim rp-address 192.168.0.3 group-address 224.0.0.0/4
```

29

Multiple Spanning Tree Protocol (MSTP)

Multiple spanning tree protocol (MSTP) is supported on the S6000 platform.

Protocol Overview

MSTP — specified in IEEE 802.1Q-2003 — is a rapid spanning tree protocol (RSTP)-based spanning tree variation that improves on per-VLAN spanning tree plus (PVST+). MSTP allows multiple spanning tree instances and allows you to map many VLANs to one spanning tree instance to reduce the total number of required instances.

In contrast, PVST+ allows a spanning tree instance for each VLAN. This 1:1 approach is not suitable if you have many VLANs, because each spanning tree instance costs bandwidth and processing resources.

In the following illustration, three VLANs are mapped to two multiple spanning tree instances (MSTI). VLAN 100 traffic takes a different path than VLAN 200 and 300 traffic. The behavior demonstrates how you can use MSTP to achieve load balancing.

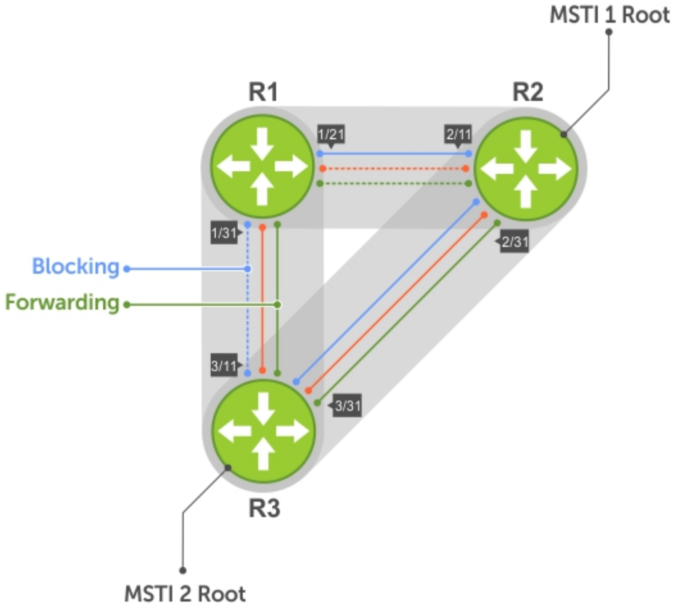


Figure 85. MSTP with Three VLANs Mapped to TWO Spanning Tree Instances

Spanning Tree Variations

The Dell Networking OS supports four variations of spanning tree, as shown in the following table.

Table 47. Spanning Tree Variations

Dell Networking Term	IEEE Specification
Spanning Tree Protocol (STP)	802 .1d
Rapid Spanning Tree Protocol (RSTP)	802 .1w
Multiple Spanning Tree Protocol (MSTP)	802 .1s
Per-VLAN Spanning Tree Plus (PVST+)	Third Party

Implementation Information

The following describes the MSTP implementation information.

- The Dell Networking OS MSTP implementation is based on IEEE 802.1Q-2003 and interoperates only with bridges that also use this standard implementation.
- MSTP is compatible with STP and RSTP.
- Dell Networking OS supports only one MSTP region.
- When you enable MSTP, all ports in Layer 2 mode participate in MSTP.

Configure Multiple Spanning Tree Protocol

Configuring multiple spanning tree is a four-step process.

1. Configure interfaces for Layer 2.
2. Place the interfaces in VLANs.
3. Enable the multiple spanning tree protocol.
4. Create multiple spanning tree instances and map VLANs to them.

Related Configuration Tasks

The following are the related configuration tasks for MSTP.

- Creating Multiple Spanning Tree Instances
- Adding and Removing Interfaces
- Influencing MSTP Root Selection
- Interoperate with Non-Dell Networking OS Bridges
- Changing the Region Name or Revision
- Modifying Global Parameters
- Modifying the Interface Parameters
- Configuring an EdgePort
- Flush MAC Addresses after a Topology Change
- Debugging and Verifying MSTP Configurations
- Prevent Network Disruptions with BPDU Guard

- Enabling SNMP Traps for Root Elections and Topology Changes
- Configuring Spanning Trees as Hitless

Enable Multiple Spanning Tree Globally

MSTP is not enabled by default. To enable MSTP globally, use the following commands.
When you enable MSTP, all physical, VLAN, and port-channel interfaces that are enabled and in Layer 2 mode are automatically part of the MSTI 0.

- Within an MSTI, only one path from any bridge to any other bridge is enabled.
- Bridges block a redundant path by disabling one of the link ports.

1. Enter PROTOCOL MSTP mode.

   CONFIGURATION mode

   ```
   protocol spanning-tree mstp
   ```

2. Enable MSTP.

   PROTOCOL MSTP mode

   ```
   no disable
   ```

Example of Verifying MSTP is Enabled

To verify that MSTP is enabled, use the `show config` command in PROTOCOL MSTP mode.

```
Dell(conf)#protocol spanning-tree mstp
Dell(config-mstp)#show config
!
protocol spanning-tree mstp
  no disable
Dell#
```

Adding and Removing Interfaces

To add and remove interfaces, use the following commands.
To add an interface to the MSTP topology, configure it for Layer 2 and add it to a VLAN.

If you previously disabled MSTP on the interface using the `no spanning-tree 0` command, to enable MSTP, use the following command.

- `spanning-tree 0`

To remove an interface from the MSTP topology, use the `no spanning-tree 0` command.

Creating Multiple Spanning Tree Instances

To create multiple spanning tree instances, use the following command.
A single MSTI provides no more benefit than RSTP. To take full advantage of MSTP, create multiple MSTIs and map VLANs to them.

- Create an MSTI.

  PROTOCOL MSTP mode

  ```
  msti
  ```

Specify the keyword `vlan` then the VLANs that you want to participate in the MSTI.

Examples of Configuring and Viewing MSTI

The following examples shows the `msti` command.

```
Dell(conf)#protocol spanning-tree mstp
Dell(conf-mstp)#msti 1 vlan 100
Dell(conf-mstp)#msti 2 vlan 200-300
Dell(conf-mstp)#show config
!
protocol spanning-tree mstp
  no disable
  MSTI 1 VLAN 100
  MSTI 2 VLAN 200-300
```

All bridges in the MSTP region must have the same VLAN-to-instance mapping.

To view which instance a VLAN is mapped to, use the `show spanning-tree mst vlan` command from EXEC Privilege mode.

```
Dell(conf-mstp)#name my-mstp-region
Dell(conf-mstp)#exit
Dell(conf)#do show spanning-tree mst config
MST region name: my-mstp-region
Revision: 0
MSTI VID
1 100
2 200-300
```

To view the forwarding/discarding state of the ports participating in an MSTI, use the `show spanning-tree msti` command from EXEC Privilege mode.

```
Dell#show spanning-tree msti 1
MSTI 1 VLANs mapped 100

Root Identifier has priority 32768, Address 0001.e806.953e
Root Bridge hello time 2, max age 20, forward delay 15, max hops 19
Bridge Identifier has priority 32768, Address 0001.e80d.b6d6
Configured hello time 2, max age 20, forward delay 15, max hops 20
Current root has priority 32768, Address 0001.e806.953e
Number of topology changes 2, last change occured 1d2h ago on Gi 1/21

Port 374 (GigabitEthernet 1/21) is root Forwarding
Port path cost 20000, Port priority 128, Port Identifier 128.374
Designated root has priority 32768, address 0001.e806.953e
Designated bridge has priority 32768, address 0001.e806.953e
Designated port id is 128.374, designated path cost 20000
Number of transitions to forwarding state 1
BPDU (MRecords): sent 93671, received 46843
The port is not in the Edge port mode

Port 384 (GigabitEthernet 1/31) is alternate Discarding
Port path cost 20000, Port priority 128, Port Identifier 128.384
Designated root has priority 32768, address 0001.e806.953e
Designated bridge has priority 32768, address 0001.e809.c24a
Designated port id is 128.384, designated path cost 20000
Number of transitions to forwarding state 1
BPDU (MRecords): sent 39291, received 7547
The port is not in the Edge port mode
```

Influencing MSTP Root Selection

MSTP determines the root bridge, but you can assign one bridge a lower priority to increase the probability that it becomes the root bridge.
To change the bridge priority, use the following command.

- Assign a number as the bridge priority.

  PROTOCOL MSTP mode

  `msti *instance* bridge-priority *priority*`

  A lower number increases the probability that the bridge becomes the root bridge.

  The range is from 0 to 61440, in increments of 4096.

  The default is **32768**.

Example of Assigning and Verifying the Root Bridge Priority

By default, the simple configuration shown previously yields the same forwarding path for both MSTIs. The following example shows how R3 is assigned bridge priority 0 for MSTI 2, which elects a different root bridge than MSTI 2.

To view the bridge priority, use the `show config` command from PROTOCOL MSTP mode.

```
R3(conf-mstp)#msti 2 bridge-priority 0
1d2h51m: %RPM0-P:RP2 %SPANMGR-5-STP_ROOT_CHANGE: MSTP root changed for instance
2. My
Bridge ID: 0:0001.e809.c24a Old Root: 32768:0001.e806.953e New Root:
0:0001.e809.c24a

R3(conf-mstp)#show config
!
protocol spanning-tree mstp
  no disable
  MSTI 1 VLAN 100
  MSTI 2 VLAN 200,300
  MSTI 2 bridge-priority 0
```

Interoperate with Non-Dell Networking OS Bridges

Dell Networking OS supports only one MSTP region.

A region is a combination of three unique qualities:

- **Name** is a mnemonic string you assign to the region. The default region name on Dell Networking OS is **null**.
- **Revision** is a 2-byte number. The default revision number on Dell Networking OS is **0**.
- VLAN-to-instance mapping is the placement of a VLAN in an MSTI.

For a bridge to be in the same MSTP region as another, all three of these qualities must match exactly. The default values for name and revision match on all Dell Networking OS equipment. If you have non-Dell Networking OS equipment that participates in MSTP, ensure these values match on all the equipment.

NOTE: Some non-Dell Networking OS equipment may implement a non-null default region name. SFTOS, for example, uses the Bridge ID, while others may use a MAC address.

Changing the Region Name or Revision

To change the region name or revision, use the following commands.

- Change the region name.

  PROTOCOL MSTP mode

  ```
  name name
  ```

- Change the region revision number.

  PROTOCOL MSTP mode

  ```
  revision number
  ```

Example of the `name` Command

To view the current region name and revision, use the `show spanning-tree mst configuration` command from EXEC Privilege mode.

```
Dell(conf-mstp)#name my-mstp-region
Dell(conf-mstp)#exit
Dell(conf)#do show spanning-tree mst config
MST region name: my-mstp-region
Revision: 0
MSTI   VID
  1    100
  2    200-300
```

Modifying Global Parameters

The root bridge sets the values for forward-delay, hello-time, max-age, and max-hops and overwrites the values set on other MSTP bridges.

- **Forward-delay** — the amount of time an interface waits in the Listening state and the Learning state before it transitions to the Forwarding state.
- **Hello-time** — the time interval in which the bridge sends MSTP bridge protocol data units (BPDUs).
- **Max-age** — the length of time the bridge maintains configuration information before it refreshes that information by recomputing the MST topology.
- **Max-hops** — the maximum number of hops a BPDU can travel before a receiving switch discards it.

NOTE: Dell Networking recommends that only experienced network administrators change MSTP parameters. Poorly planned modification of MSTP parameters can negatively affect network performance.

To change the MSTP parameters, use the following commands on the root bridge.

1. Change the forward-delay parameter.

   PROTOCOL MSTP mode

   ```
   forward-delay seconds
   ```

   The range is from 4 to 30.

The default is **15 seconds**.

2. Change the hello-time parameter.

   PROTOCOL MSTP mode

   ```
   hello-time seconds
   ```

   NOTE: With large configurations (especially those configurations with more ports) Dell Networking recommends increasing the hello-time.

   The range is from 1 to 10.

   The default is **2 seconds**.

3. Change the max-age parameter.

   PROTOCOL MSTP mode

   ```
   max-age seconds
   ```

   The range is from 6 to 40.

   The default is **20 seconds**.

4. Change the max-hops parameter.

   PROTOCOL MSTP mode

   ```
   max-hops number
   ```

   The range is from 1 to 40.

   The default is **20**.

Example of the `forward-delay` Parameter

To view the current values for MSTP parameters, use the `show running-config spanning-tree mstp` command from EXEC privilege mode.

```
Dell(conf-mstp)#forward-delay 16
Dell(conf-mstp)#exit
Dell(conf)#do show running-config spanning-tree mstp
!
protocol spanning-tree mstp
no disable
name my-mstp-region
MSTI 1 VLAN 100
MSTI 2 VLAN 200-300
forward-delay 16
MSTI 2 bridge-priority 4096
Dell(conf)#
```

Modifying the Interface Parameters

You can adjust two interface parameters to increase or decrease the probability that a port becomes a forwarding port.

- **Port cost** is a value that is based on the interface type. The greater the port cost, the less likely the port is selected to be a forwarding port.

- **Port priority** influences the likelihood that a port is selected to be a forwarding port in case that several ports have the same port cost.

The following lists the default values for port cost by interface.

Table 48. Default Values for Port Costs by Interface

Port Cost	Default Value
100-Mb/s Ethernet interfaces	200000
1-Gigabit Ethernet interfaces	20000
10-Gigabit Ethernet interfaces	2000
Port Channel with 100 Mb/s Ethernet interfaces	180000
Port Channel with 1-Gigabit Ethernet interfaces	18000
Port Channel with 10-Gigabit Ethernet interfaces	1800

To change the port cost or priority of an interface, use the following commands.

1. Change the port cost of an interface.
   INTERFACE mode

   ```
   spanning-tree msti number cost cost
   ```

   The range is from 0 to 200000.

   For the default, refer to the default values shown in the table..
2. Change the port priority of an interface.
   INTERFACE mode

   ```
   spanning-tree msti number priority priority
   ```

   The range is from 0 to 240, in increments of 16.

   The default is **128**.

To view the current values for these interface parameters, use the `show config` command from INTERFACE mode.

Configuring an EdgePort

The EdgePort feature enables interfaces to begin forwarding traffic approximately 30 seconds sooner. In this mode, an interface forwards frames by default until it receives a BPDU that indicates that it should behave otherwise; it does not go through the Learning and Listening states. The `bpduguard shutdown-on-violation` option causes the interface hardware to be shut down when it receives a BPDU. When you implement only `bpduguard`, although the interface is placed in an Error Disabled state when receiving the BPDU, the physical interface remains up and spanning-tree drops packets in the hardware after a BPDU violation. BPDUs are dropped in the software after receiving the BPDU violation. This feature is the same as PortFast mode in spanning tree.

⚠ **CAUTION: Configure EdgePort only on links connecting to an end station. EdgePort can cause loops if you enable it on an interface connected to a network.**

To enable EdgePort on an interface, use the following command.

- Enable EdgePort on an interface.

  INTERFACE mode

  `spanning-tree mstp edge-port [bpduguard | shutdown-on-violation]`

  Dell Networking OS Behavior: Regarding `bpduguard shutdown-on-violation` behavior:

  - If the interface to be shut down is a port channel, all the member ports are disabled in the hardware.
  - When you add a physical port to a port channel already in the Error Disable state, the new member port is also disabled in the hardware.
  - When you remove a physical port from a port channel in the Error Disable state, the error disabled state is cleared on this physical port (the physical port is enabled in the hardware).
  - The `reset linecard` command does not clear the Error Disabled state of the port or the Hardware Disabled state. The interface continues to be disabled in the hardware.
  - You can clear the Error Disabled state with any of the following methods:
    - Use the `shutdown` command on the interface.
    - Disable the `shutdown-on-violation` command on the interface (using the `no spanning-tree stp-id portfast [bpduguard | [shutdown-on-violation]]` command).
    - Disable spanning tree on the interface (using the `no spanning-tree` command in INTERFACE mode).
    - Disabling global spanning tree (using the `no spanning-tree` command in CONFIGURATION mode).

Example of Enabling an EdgePort on an Interface

To verify that EdgePort is enabled, use the `show config` command from INTERFACE mode.

```
Dell(conf-if-gi-3/41)#spanning-tree mstp edge-port
Dell(conf-if-gi-3/41)#show config
!
interface GigabitEthernet 3/41
  no ip address
  switchport
  spanning-tree mstp edge-port
  spanning-tree MSTI 1 priority 144
  no shutdown
Dell(conf-if-gi-3/41)#
```

Flush MAC Addresses after a Topology Change

Dell Networking OS has an optimized MAC address flush mechanism for RSTP, MSTP, and PVST+ that flushes addresses only when necessary, which allows for faster convergence during topology changes.

However, you may activate the flushing mechanism defined by 802.1Q-2003 using the `tc-flush-standard` command, which flushes MAC addresses after every topology change notification.

To view the enable status of this feature, use the `show running-config spanning-tree mstp` command from EXEC Privilege mode.

MSTP Sample Configurations

The running-configurations support the topology shown in the following illustration.

The configurations are from Dell Networking OS systems.

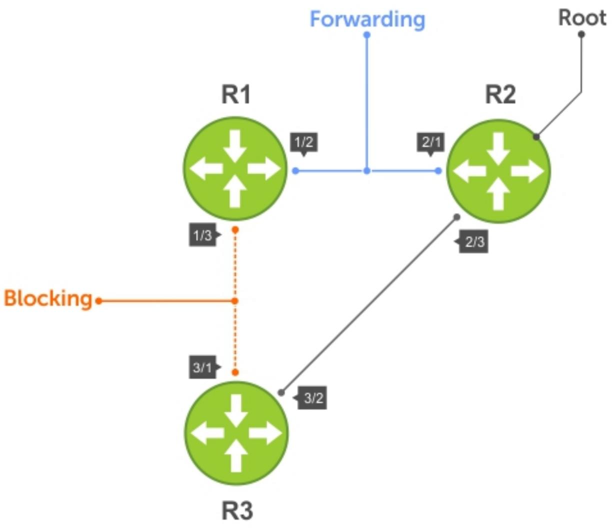


Figure 86. MSTP with Three VLANs Mapped to Two Spanning Tree Instances

Router 1 Running-Configuration

This example uses the following steps:

1. Enable MSTP globally and set the region name and revision map MSTP instances to the VLANs.
2. Assign Layer-2 interfaces to the MSTP topology.
3. Create VLANs mapped to MSTP instances tag interfaces to the VLANs.

(Step 1)

```
protocol spanning-tree mstp
  no disable
  name Tahiti
  revision 123
  MSTI 1 VLAN 100
  MSTI 2 VLAN 200,300
!
```

(Step 2)

```
interface GigabitEthernet 1/21
  no ip address
  switchport
  no shutdown
!
interface GigabitEthernet 1/31
  no ip address
  switchport
  no shutdown
!
```

(Step 3)

```
interface Vlan 100
  no ip address
  tagged GigabitEthernet 1/21,31
  no shutdown
!
interface Vlan 200
  no ip address
  tagged GigabitEthernet 1/21,31
  no shutdown
!
interface Vlan 300
  no ip address
  tagged GigabitEthernet 1/21,31
  no shutdown
```

Router 2 Running-Configuration

This example uses the following steps:

1. Enable MSTP globally and set the region name and revision map MSTP instances to the VLANs.
2. Assign Layer-2 interfaces to the MSTP topology.
3. Create VLANs mapped to MSTP instances tag interfaces to the VLANs.

(Step 1)

```
protocol spanning-tree mstp
  no disable
  name Tahiti
  revision 123
  MSTI 1 VLAN 100
  MSTI 2 VLAN 200,300
!
```

(Step 2)

```
interface GigabitEthernet 2/11
  no ip address
  switchport
  no shutdown
!
interface GigabitEthernet 2/31
  no ip address
  switchport
  no shutdown
!
```

(Step 3)

```
interface Vlan 100
  no ip address
  tagged GigabitEthernet 2/11,31
 no shutdown
!
interface Vlan 200
  no ip address
  tagged GigabitEthernet 2/11,31
  no shutdown
!
interface Vlan 300
  no ip address
  tagged GigabitEthernet 2/11,31
  no shutdown
```

Router 3 Running-Configuration

This example uses the following steps:

1. Enable MSTP globally and set the region name and revision map MSTP instances to the VLANs.
2. Assign Layer-2 interfaces to the MSTP topology.
3. Create VLANs mapped to MSTP instances tag interfaces to the VLANs.

(Step 1)

```
protocol spanning-tree mstp
  no disable
  name Tahiti
  revision 123
  MSTI 1 VLAN 100
  MSTI 2 VLAN 200,300
!
```

(Step 2)

```
interface GigabitEthernet 3/11
  no ip address
  switchport
  no shutdown
!
interface GigabitEthernet 3/21
  no ip address
  switchport
  no shutdown
!
```

(Step 3)

```
interface Vlan 100
  no ip address
  tagged GigabitEthernet 3/11,21
  no shutdown
!
interface Vlan 200
  no ip address
  tagged GigabitEthernet 3/11,21
  no shutdown
!
interface Vlan 300
  no ip address
  tagged GigabitEthernet 3/11,21
  no shutdown
```

SFTOS Example Running-Configuration

This example uses the following steps:

1. Enable MSTP globally and set the region name and revision map MSTP instances to the VLANs.
2. Assign Layer-2 interfaces to the MSTP topology.
3. Create VLANs mapped to MSTP instances tag interfaces to the VLANs.

(Step 1)

```
spanning-tree
spanning-tree configuration name Tahiti
spanning-tree configuration revision 123
spanning-tree MSTi instance 1
spanning-tree MSTi vlan 1 100
spanning-tree MSTi instance 2
spanning-tree MSTi vlan 2 200
spanning-tree MSTi vlan 2 300
```

(Step 2)

```
interface 1/0/31
  no shutdown
  spanning-tree port mode enable
  switchport protected 0
exit

interface 1/0/32
  no shutdown
  spanning-tree port mode enable
  switchport protected 0
exit
```

(Step 3)

```
interface vlan 100
  tagged 1/0/31
  tagged 1/0/32
exit

interface vlan 200
  tagged 1/0/31
  tagged 1/0/32
exit

interface vlan 300
  tagged 1/0/31
  tagged 1/0/32
exit
```

Debugging and Verifying MSTP Configurations

To debut and verify MSTP configuration, use the following commands.

- Display BPDUs.

  EXEC Privilege mode

  `debug spanning-tree mstp bpdu`
- Display MSTP-triggered topology change messages.

  `debug spanning-tree mstp events`

Examples of Viewing MSTP Configurations

To ensure all the necessary parameters match (region name, region version, and VLAN to instance mapping), examine your individual routers.

To show various portions of the MSTP configuration, use the `show spanning-tree mst` commands.

To view the overall MSTP configuration on the router, use the `show running-configuration spanning-tree mstp` in EXEC Privilege mode.

To monitor and verify that the MSTP configuration is connected and communicating as desired, use the `debug spanning-tree mstp bpdu` command.

Key items to look for in the debug report include:

- MSTP flags indicate communication received from the same region.
  - As shown in the following, the MSTP routers are located in the same region.
  - Does the debug log indicate that packets are coming from a "Different Region"? If so, one of the key parameters is not matching.
- MSTP Region Name and Revision.
  - The configured name and revisions must be identical among all the routers.
  - Is the Region name blank? That may mean that a name was configured on one router and but was not configured or was configured differently on another router (spelling and capitalization counts).
- MSTP Instances.
  - To verify the VLAN to MSTP instance mapping, use the `show` commands.

  - Are there “extra” MSTP instances in the Sending or Received logs? This may mean that an additional MSTP instance was configured on one router but not the others.

The following example shows the `show run spanning-tree mstp` command.

```
Dell#show run spanning-tree mstp
!
protocol spanning-tree mstp
name Tahiti
revision 123
MSTI 1 VLAN 100
MSTI 2 VLAN 200,300
```

The following example shows viewing the debug log of a successful MSTP configuration.

```
Dell#debug spanning-tree mstp bpdu
MSTP debug bpdu is ON
Dell#
4w0d4h : MSTP: Sending BPDU on Gi 2/21 :
ProtId: 0, Ver: 3, Bpdu Type: MSTP, Flags 0x6e
CIST Root Bridge Id: 32768:0001.e806.953e, Ext Path Cost: 0
Regional Bridge Id: 32768:0001.e806.953e, CIST Port Id: 128:470
Msg Age: 0, Max Age: 20, Hello: 2, Fwd Delay: 15, Ver1 Len: 0, Ver3 Len: 96
Name: Tahiti, Rev: 123, Int Root Path Cost: 0
Rem Hops: 20, Bridge Id: 32768:0001.e806.953e
4w0d4h : INST 1: Flags: 0x6e, Reg Root: 32768:0001.e806.953e, Int Root Cost: 0
    Brg/Port Prio: 32768/128, Rem Hops: 20
INST 2: Flags: 0x6e, Reg Root: 32768:0001.e806.953e, Int Root Cost: 0
    Brg/Port Prio: 32768/128, Rem Hops: 20

4w0d4h : MSTP: Received BPDU on Gi 2/21 :
ProtId: 0, Ver: 3, Bpdu Type: MSTP, Flags 0x78 (Indicates MSTP routers are in the [single]
region.)
CIST Root Bridge Id: 32768:0001.e806.953e, Ext Path Cost: 0
Regional Bridge Id: 32768:0001.e806.953e, CIST Port Id: 128:470
Msg Age: 0, Max Age: 20, Hello: 2, Fwd Delay: 15, Ver1 Len: 0, Ver3 Len: 96
Name: Tahiti, Rev: 123 (MSTP region name and revision), Int Root Path Cost: 0
Rem Hops: 19, Bridge Id: 32768:0001.e8d5.cbbd
4w0d4h : INST 1 (MSTP Instance): Flags: 0x78, Reg Root: 32768:0001.e806.953e, Int
Root Cost: 0
    Brg/Port Prio: 32768/128, Rem Hops: 19
INST 2 (MSTP Instance): Flags: 0x78, Reg Root: 32768:0001.e806.953e, Int Root Cost:
0
    Brg/Port Prio: 32768/128, Rem Hops: 19
Indicates MSTP
routers are in the
(single) region
MSTP Instance
MSTP Region name
```

The following example shows viewing the debug log of an unsuccessful MSTP configuration.

```
4w0d4h : MSTP: Received BPDU on Gi 2/21 :
ProtId: 0, Ver: 3, Bpdu Type: MSTP, Flags 0x78Different Region (Indicates MSTP routers
are in different regions and are not communicating with each other.)
CIST Root Bridge Id: 32768:0001.e806.953e, Ext Path Cost: 0
Regional Bridge Id: 32768:0001.e806.953e, CIST Port Id: 128:470
Msg Age: 0, Max Age: 20, Hello: 2, Fwd Delay: 15, Ver1 Len: 0, Ver
Name: Tahiti, Rev: 123, Int Root Path Cost: 0
Rem Hops: 20, Bridge Id: 32768:0001.e8d5.cbbd
4w0d4h : INST 1: Flags: 0x70, Reg Root: 32768:0001.e8d5.cbbd, Int
    Brg/Port Prio: 32768/128, Rem Hops: 20
```

```
INST 2: Flags: 0x70, Reg Root: 32768:0001.e8d5.cbbd, Int Root Cost
    Brg/Port Prio: 32768/128, Rem Hops: 20
```

30

Multicast Features

Multicast features are supported on the S6000 platform.

NOTE: Multicast is supported on secondary IP addresses on the platform.

NOTE: Multicast routing for IPv6 is not supported.

The Dell Networking Operating System (OS) supports the following multicast protocols:

- PIM Sparse-Mode (PIM-SM)
- Internet Group Management Protocol (IGMP)
- Multicast Source Discovery Protocol (MSDP)

Enabling IP Multicast

Enable IP multicast is supported on the S6000 platform.
Prior to enabling any multicast protocols, you must enable multicast routing.

- Enable multicast routing.

  CONFIGURATION mode

  ```
  ip multicast-routing
  ```

Multicast with ECMP

Dell Networking multicast uses equal-cost multi-path (ECMP) routing to load-balance multiple streams across equal cost links.

When creating the shared-tree protocol independent multicast (PIM) uses routes from all configured routing protocols to select the best route to the rendezvous point (RP). If there are multiple, equal-cost paths, the PIM selects the route with the least number of currently running multicast streams. If multiple routes have the same number of streams, PIM selects the first equal-cost route the route table manager (RTM) returns.

In the following illustration, the receiver joins three groups. The last-hop DR initially has two equal-cost routes to the RP with no streams, so it non-deterministically selects Route 1 for the Group 1 IGMP Join message. Route 1 then has one stream associated with it, so the last-hop DR sends the Group 2 Join by Route 2. It then non-deterministically selects Route 2 for the Group 3 Join because both routes already have one multicast stream.

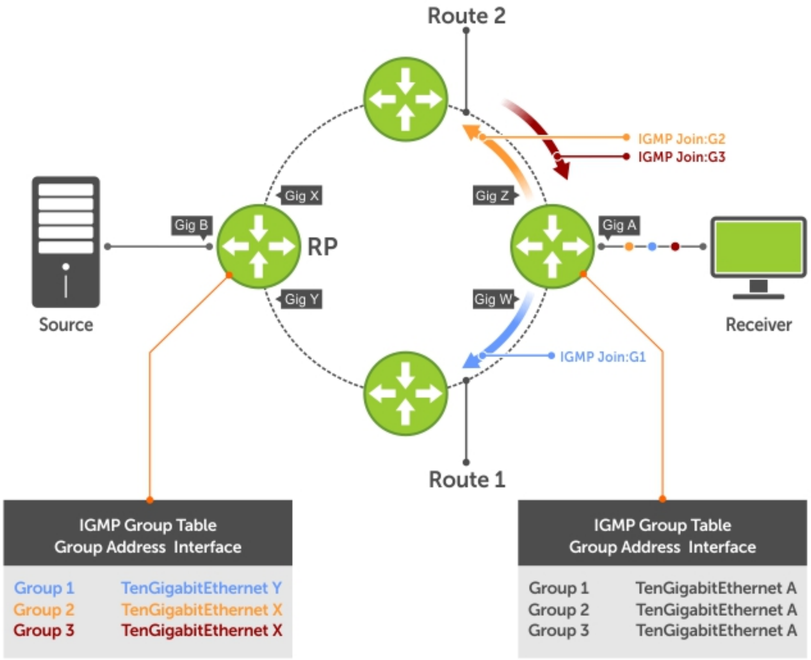

Route 2
IGMP Join:G2
IGMP Join:G3
Gig X
Gig Z
Gig B
RP
Gig A
Source
Gig Y
Gig W
Receiver
IGMP Join:G1
Route 1
IGMP Group Table
Group Address Interface
Group 1 TenGigabitEthernet Y
Group 2 TenGigabitEthernet X
Group 3 TenGigabitEthernet X
IGMP Group Table
Group Address Interface
Group 1 TenGigabitEthernet A
Group 2 TenGigabitEthernet A
Group 3 TenGigabitEthernet A

Figure 87. Multicast with ECMP

Implementation Information

Because protocol control traffic in Dell Networking OS is redirected using the MAC address, and multicast control traffic and multicast data traffic might map to the same MAC address, Dell Networking OS might forward data traffic with certain MAC addresses to the CPU in addition to control traffic.

As the upper5 bits of an IP Multicast address are dropped in the translation, 32 different multicast group IDs all map to the same Ethernet address. For example, 224.0.0.5 is a known IP address for open shortest path first (OSPF) that maps to the multicast MAC address 01:00:5e:00:00:05. However, 225.0.0.5, 226.0.0.5, and so on, map to the same multicast MAC address. The Layer 2 forwarding information base (FIB) alone cannot differentiate multicast control traffic multicast data traffic with the same address, so if you use IP address 225.0.0.5 for data traffic, both the multicast data and OSPF control traffic match the same entry and are forwarded to the CPU. Therefore, do not use well-known protocol multicast addresses for data transmission, such as the following.

Protocol	Ethernet Address
OSPF	01:00:5e:00:00:05 01:00:5e:00:00:06
RIP	01:00:5e:00:00:09
NTP	01:00:5e:00:01:01
VRRP	01:00:5e:00:00:12

Protocol	Ethernet Address
PIM-SM	01:00:5e:00:00:0d

- The Dell Networking OS implementation of MTRACE is in accordance with IETF draft *draft-fenner-traceroute-ipm*.
- Multicast is not supported on secondary IP addresses.
- Egress L3 ACL is not applied to multicast data traffic if you enable multicast routing.

First Packet Forwarding for Lossless Multicast

All initial multicast packets are forwarded to receivers to achieve lossless multicast.

In previous versions, when the Dell Networking system is an RP, all initial packets are dropped until PIM creates an (S,G) entry. When the system is an RP and a Source DR, these initial packet drops represent a loss of native data, and when the system is an RP only, the initial packets drops represent a loss of register packets.

Both scenarios might be unacceptable depending on the multicast application. When the Dell Networking system is the RP, and has receivers for a group G, it forwards all initial multicast packets for the group based on the (*,G) entry rather than discarding them until the (S,G) entry is created, making Dell Networking systems suitable for applications sensitive to multicast packet loss.

NOTE: When a source begins sending traffic, the Source DR forwards the initial packets to the RP as encapsulated registered packets. These packets are forwarded via the soft path at a maximum rate of 70 packets/second. Incoming packets beyond this rate are dropped.

Multicast Policies

Dell Networking OS offers parallel multicast features for IPv4.

- IPv4 Multicast Policies

IPv4 Multicast Policies

The following sections describe IPv4 multicast policies.

- Limiting the Number of Multicast Routes
- Preventing a Host from Joining a Group
- Rate Limiting IGMP Join Requests
- Preventing a PIM Router from Forming an Adjacency
- Preventing a Source from Registering with the RP
- Preventing a PIM Router from Processing a Join

Limiting the Number of Multicast Routes

When the total number of multicast routes on a system limit is reached, Dell Networking OS does not process any IGMP or multicast listener discovery protocol (MLD) joins to PIM — though it still processes leave messages — until the number of entries decreases below 95% of the limit.

When the limit falls below 95% after hitting the maximum, the system begins relearning route entries through IGMP, MLD, and MSDP.

- If the limit is increased after it is reached, join subsequent join requests are accepted. In this case, increase the limit by at least 10% for IGMP and MLD to resume.

- If the limit is decreased after it is reached, Dell Networking OS does not clear the existing sessions. Entries are cleared after a timeout (you may also clear entries using `clear ip mroute`).

NOTE: Dell Networking OS waits at least 30 seconds between stopping and starting IGMP join processing. You may experience this delay when manipulating the limit after it is reached.

When the multicast route limit is reached, Dell Networking OS displays the following:

```
      3w1d13h: %RPM0-P:RP2 %PIM-3-PIM_TIB_LIMIT: PIM TIB limit reached. No new
routes will
be learnt until TIB level falls below low watermark.
      3w1d13h: %RPM0-P:RP2 %PIM-3-PIM_TIB_LIMIT: PIM TIB below low watermark.
Route learning
will begin.
```

To limit the number of multicast routes, use the following command.

- Limit the total number of multicast routes on the system.

  CONFIGURATION mode

  ```
  ip multicast-limit
  ```

  The range if from 1 to 50000.

  The default is **15000**.

NOTE: The IN-L3-McastFib CAM partition is used to store multicast routes and is a separate hardware limit that exists per port-pipe. Any software-configured limit may supersede by this hardware space limitation. The opposite is also true, the CAM partition might not be exhausted at the time the system-wide route limit the `ip multicast-limit` command sets is reached.

Preventing a Host from Joining a Group

You can prevent a host from joining a particular group by blocking specific IGMP reports. Create an extended access list containing the permissible source-group pairs.

NOTE: For rules in IGMP access lists, *source* is the multicast source, not the source of the IGMP packet. For IGMPv2, use the keyword `any` for *source* (as shown in the following example), because IGMPv2 hosts do not know in advance who the source is for the group in which they are interested.

To apply the access list, use the following command.

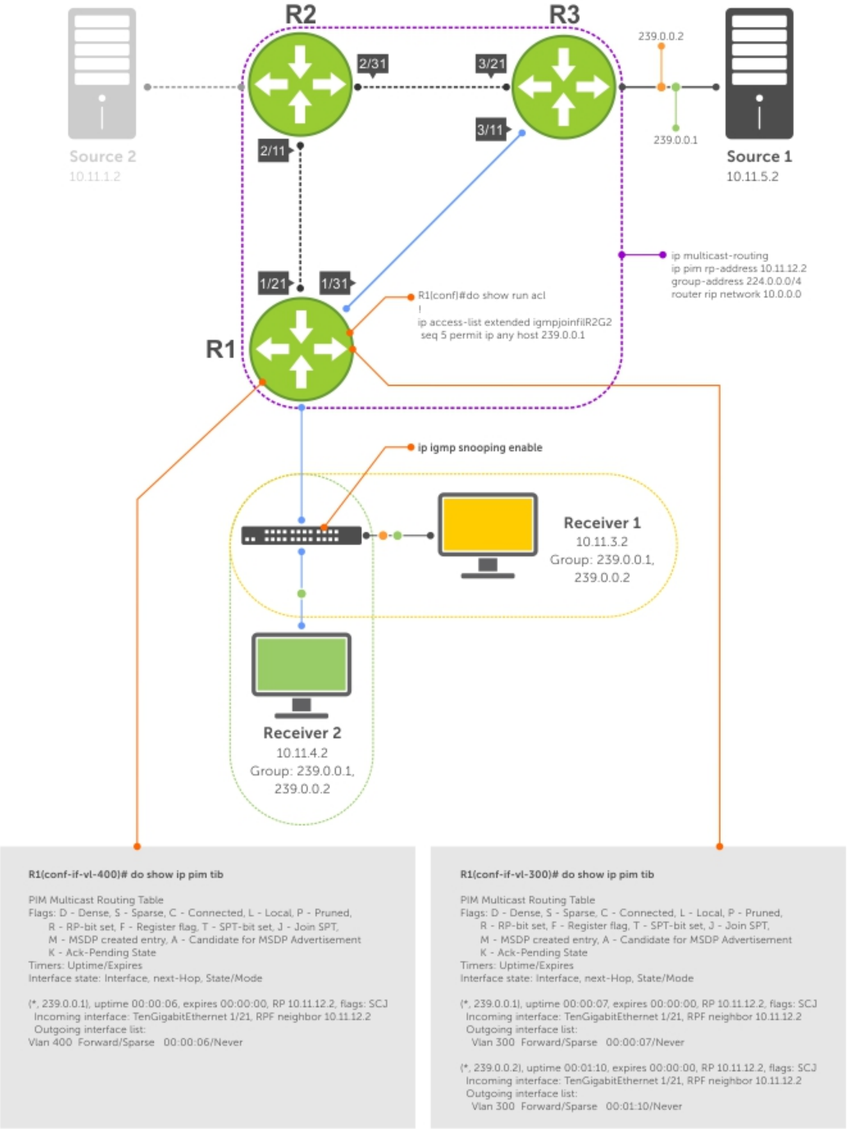
- Apply the access list.

  INTERFACE mode

  ```
  ip igmp access-group access-list-name
  ```

Dell Networking OS Behavior: Do not enter the `ip igmp access-group` command before creating the access-list. If you do, after entering your first deny rule, Dell Networking OS clears multicast routing table and re-learns all groups, even those not covered by the rules in the access-list, because there is an implicit *deny all* rule at the end of all access-lists. Therefore, configuring an IGMP join request filter in this order might result in data loss. If you must enter the `ip igmp access-group` command before creating the access-list, prevent Dell Networking OS from clearing the routing table by entering a *permit any* rule with high sequence number before you enter any other rules.

In the following example, virtual local area network (VLAN) 400 is configured with an access list to permit only IGMP reports for group 239.0.0.1. Though Receiver 2 sends a membership report for groups 239.0.0.1 and 239.0.0.2, a multicast routing table entry is created only for group 239.0.0.1. VLAN 300 has no access list limiting Receiver 1, so both IGMP reports are accepted, and two corresponding entries are created in the routing table.



```
ip multicast-routing
ip pim rp-address 10.11.12.2
group-address 224.0.0.0/4
router rip network 10.0.0.0
```

```
R1(conf)#do show run acl
!
ip access-list extended igmpjoinfilR2G2
 seq 5 permit ip any host 239.0.0.1
```

```
R1(conf-if-vl-400)# do show ip pim tib

PIM Multicast Routing Table
Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned,
       R - RP-bit set, F - Register flag, T - SPT-bit set, J - Join SPT,
       M - MSDP created entry, A - Candidate for MSDP Advertisement
       K - Ack-Pending State
Timers: Uptime/Expires
Interface state: Interface, next-Hop, State/Mode

(*, 239.0.0.1), uptime 00:00:06, expires 00:00:00, RP 10.11.12.2, flags: SCJ
  Incoming interface: TenGigabitEthernet 1/21, RPF neighbor 10.11.12.2
  Outgoing interface list:
Vlan 400  Forward/Sparse   00:00:06/Never
```

```
R1(conf-if-vl-300)# do show ip pim tib

PIM Multicast Routing Table
Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned,
       R - RP-bit set, F - Register flag, T - SPT-bit set, J - Join SPT,
       M - MSDP created entry, A - Candidate for MSDP Advertisement
       K - Ack-Pending State
Timers: Uptime/Expires
Interface state: Interface, next-Hop, State/Mode

(*, 239.0.0.1), uptime 00:00:07, expires 00:00:00, RP 10.11.12.2, flags: SCJ
  Incoming interface: TenGigabitEthernet 1/21, RPF neighbor 10.11.12.2
  Outgoing interface list:
    Vlan 300  Forward/Sparse   00:00:07/Never

(*, 239.0.0.2), uptime 00:01:10, expires 00:00:00, RP 10.11.12.2, flags: SCJ
  Incoming interface: TenGigabitEthernet 1/21, RPF neighbor 10.11.12.2
  Outgoing interface list:
    Vlan 300  Forward/Sparse   00:01:10/Never
```

Figure 88. Preventing a Host from Joining a Group

Table 49. Preventing a Host from Joining a Group — Description

Location	Description
1/21	• Interface GigabitEthernet 1/21 • ip pim sparse-mode • ip address 10.11.12.1/24

Location	Description
	• no shutdown
1/31	• Interface GigabitEthernet 1/31 • ip pim sparse-mode • ip address 10.11.13.1/24 • no shutdown
2/1	• Interface GigabitEthernet 2/1 • ip pim sparse-mode • ip address 10.11.1.1/24 • no shutdown
2/11	• Interface GigabitEthernet 2/11 • ip pim sparse-mode • ip address 10.11.12.2/24 • no shutdown
2/31	• Interface GigabitEthernet 2/31 • ip pim sparse-mode • ip address 10.11.23.1/24 • no shutdown
3/1	• Interface GigabitEthernet 3/1 • ip pim sparse-mode • ip address 10.11.5.1/24 • no shutdown
3/11	• Interface GigabitEthernet 3/11 • ip pim sparse-mode • ip address 10.11.13.2/24 • no shutdown
3/21	• Interface GigabitEthernet 3/21 • ip pim sparse-mode • ip address 10.11.23.2/24 • no shutdown
Receiver 1	• Interface VLAN 300 • ip pim sparse-mode • ip address 10.11.3.1/24 • untagged GigabitEthernet 1/1 • no shutdown
Receiver 2	• Interface VLAN 400 • ip pim sparse-mode • ip address 10.11.4.1/24 • untagged GigabitEthernet 1/2

Location	Description
	- **ip igmp access-group igmpjoinfilR2G2** - no shutdown

Rate Limiting IGMP Join Requests

If you expect a burst of IGMP Joins, protect the IGMP process from overload by limiting that rate at which new groups can be joined.
Hosts whose IGMP requests are denied will use the retry mechanism built-in to IGMP so that they're membership is delayed rather than permanently denied.

- Limit the rate at which new groups can be joined.

  INTERFACE mode

  ```
  ip igmp group-join-limit
  ```

To view the enable status of this feature, use the `show ip igmp interface` command from EXEC Privilege mode.

Preventing a PIM Router from Forming an Adjacency

To prevent a router from participating in PIM (for example, to configure stub multicast routing), use the following command.

- Prevent a router from participating in protocol independent multicast (PIM).

  INTERFACE mode

  ```
  ip pim neighbor-filter
  ```

Setting a Threshold for Switching to the SPT

The functionality to specify a threshold for switchover to the shortest path trees (SPTs) is available on theS6000 platform. After a receiver receives traffic from the RP, PM-SM switches to SPT to forward multicast traffic. Every multicast group has an RP and a unidirectional shared tree (group-specific shared tree).

The SPT-Threshold is zero, which means that the last-hop designated router (DR) joins the shortest path tree (SPT) to the source upon receiving the first multicast packet.

Initially, a single PIM-SM tree called a shared tree to distribute traffic. It is called shared because all traffic for the group, regardless of the source, or the location of the source, must pass through the RP. The shared tree is unidirectional; that is, all multicast traffic flows only from the RP to the receivers. Once a receiver receives traffic from the RP, PM-SM switches to SPT to forward multicast traffic, which connects the receiver directly to the source.

You can configure PIM to switch over to the SPT when the router receives multicast packets at or beyond a specified rate.

Table 50. Configuring PIM to Switch Over to the SPT

IPv4	Configure PIM to switch over to the SPT when the multicast packet rate is at or beyond a	ip pim spt-threshold {value | infinity}	CONFIGURATION

	specified rate. The keyword infinity directs PIM to never switch to the SPT.	Default: 10 kbps	
IPv6	Configure PIM to switch over to the SPT when the multicast packet rate is at or beyond a specified rate. The keyword infinity directs PIM to never switch to the SPT.	ip pim spt-threshold {value | infinity} Default: 10 kbps	CONFIGURATION

Preventing a Source from Registering with the RP

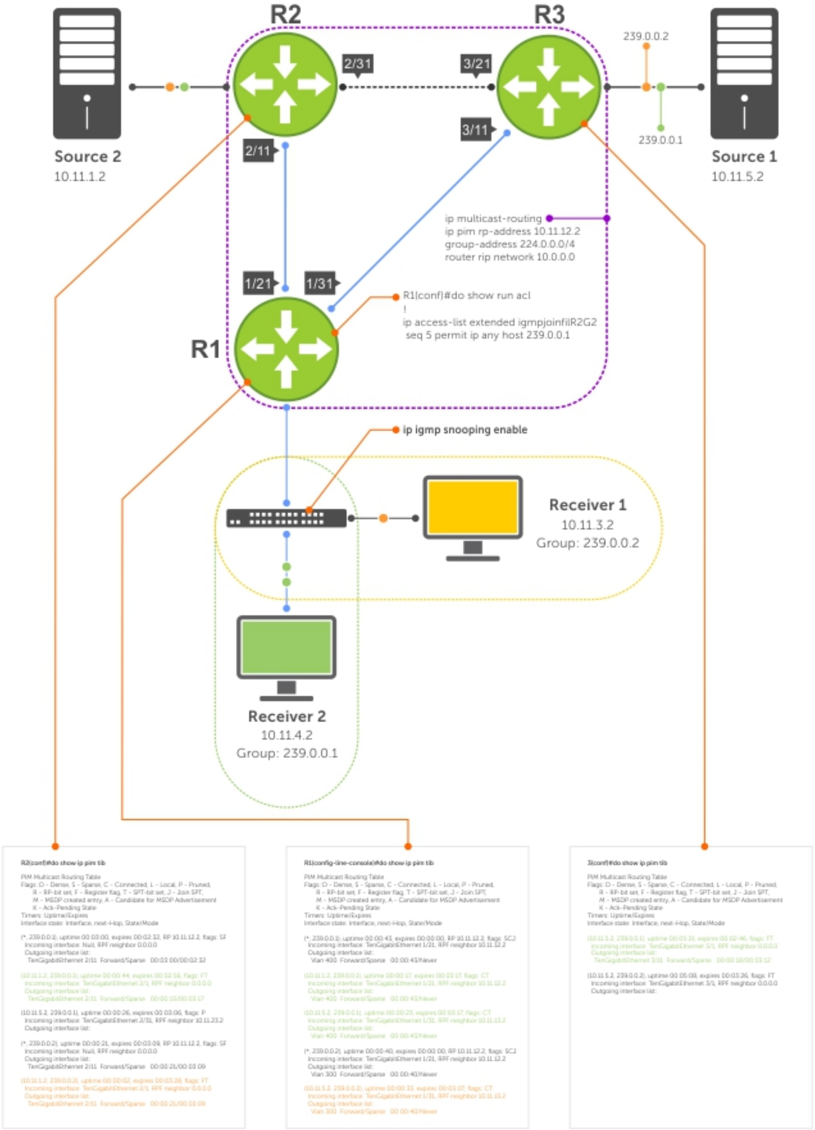
To prevent the PIM source DR from sending register packets to RP for the specified multicast source and group, use the following command. If the source DR never sends register packets to the RP, no hosts can ever discover the source and create a shortest path tree (SPT) to it.

- Prevent a source from transmitting to a particular group.

  CONFIGURATION mode

  ```
  ip pim register-filter
  ```

In the following example, Source 1 and Source 2 are both transmitting packets for groups 239.0.0.1 and 239.0.0.2. R3 has a PIM register filter that only permits packets destined for group 239.0.0.2. An entry is created for group 239.0.0.1 in the routing table, but no outgoing interfaces are listed. R2 has no filter, so it is allowed to forward both groups. As a result, Receiver 1 receives only one transmission, while Receiver 2 receives duplicate transmissions.


R2
R3
2/31
3/21
3/11
2/11
1/21
1/31
R1
Source 2
10.11.1.2
Source 1
10.11.5.2
239.0.0.2
239.0.0.1
Receiver 1
10.11.3.2
Group: 239.0.0.2
Receiver 2
10.11.4.2
Group: 239.0.0.1

```
ip multicast-routing
ip pim rp-address 10.11.12.2
group-address 224.0.0.0/4
router rip network 10.0.0.0
```

```
R1(conf)#do show run acl
!
ip access-list extended igmpjoinfilR2G2
 seq 5 permit ip any host 239.0.0.1
```

```
ip igmp snooping enable
```

Figure 89. Preventing a Source from Transmitting to a Group

Table 51. Preventing a Source from Transmitting to a Group — Description

Location	Description
1/21	• Interface GigabitEthernet 1/21 • ip pim sparse-mode • ip address 10.11.12.1/24 • no shutdown
1/31	• Interface GigabitEthernet 1/31 • ip pim sparse-mode • ip address 10.11.13.1/24

Location	Description
	• no shutdown
2/1	• Interface GigabitEthernet 2/1 • ip pim sparse-mode • ip address 10.11.1.1/24 • no shutdown
2/11	• Interface GigabitEthernet 2/11 • ip pim sparse-mode • ip address 10.11.12.2/24 • no shutdown
2/31	• Interface GigabitEthernet 2/31 • ip pim sparse-mode • ip address 10.11.23.1/24 • no shutdown
3/1	• Interface GigabitEthernet 3/1 • ip pim sparse-mode • ip address 10.11.5.1/24 • no shutdown
3/11	• Interface GigabitEthernet 3/11 • ip pim sparse-mode • ip address 10.11.13.2/24 • no shutdown
3/21	• Interface GigabitEthernet 3/21 • ip pim sparse-mode • ip address 10.11.23.2/24 • no shutdown
Receiver 1	• Interface VLAN 300 • ip pim sparse-mode • ip address 10.11.3.1/24 • untagged GigabitEthernet 1/1 • no shutdown
Receiver 2	• Interface VLAN 400 • ip pim sparse-mode • ip address 10.11.4.1/24 • untagged GigabitEthernet 1/2 • no shutdown

Preventing a PIM Router from Processing a Join

To permit or deny PIM Join/Prune messages on an interface using an extended IP access list, use the following command.

NOTE: Dell Networking recommends not using the `ip pim join-filter` command on an interface between a source and the RP router. Using this command in this scenario could cause problems with the PIM-SM source registration process resulting in excessive traffic being sent to the CPU of both the RP and PIM DR of the source.

Excessive traffic is generated when the join process from the RP back to the source is blocked due to a new source group being permitted in the join-filter. This results in the new source becoming stuck in registering on the DR and the continuous generation of UDP-encapsulated registration messages between the DR and RP routers which are being sent to the CPU.

- Prevent the PIM SM router from creating state based on multicast source and/ or group.

  ```
  ip pim join-filter
  ```

31

Open Shortest Path First (OSPFv2 and OSPFv3)

Open shortest path first (OSPFv2 for IPv4) and OSPF version 3 (OSPF for IPv6) are supported on the S6000 platform.

This chapter provides a general description of OSPFv2 (OSPF for IPv4) and OSPFv3 (OSPF for IPv6) as supported in the Dell Networking Operating System (OS).

NOTE: The fundamental mechanisms of OSPF (flooding, DR election, area support, SPF calculations, and so on) are the same between OSPFv2 and OSPFv3. This chapter identifies and clarifies the differences between the two versions of OSPF. Except where identified, the information in this chapter applies to both protocol versions.

OSPF protocol standards are listed in the Standards Compliance chapter.

Protocol Overview

OSPF routing is a link-state routing protocol that calls for the sending of link-state advertisements (LSAs) to all other routers within the same autonomous system (AS) areas.

Information on attached interfaces, metrics used, and other variables is included in OSPF LSAs. As OSPF routers accumulate link-state information, they use the shortest path first (SPF) algorithm to calculate the shortest path to each node.

OSPF routers initially exchange HELLO messages to set up adjacencies with neighbor routers. The HELLO process is used to establish adjacencies between routers of the AS. It is not required that every router within the AS areas establish adjacencies. If two routers on the same subnet agree to become neighbors through the HELLO process, they begin to exchange network topology information in the form of LSAs.

In OSPFv2 neighbors on broadcast and NBMA links are identified by their interface addresses, while neighbors on other types of links are identified by RID.

Autonomous System (AS) Areas

OSPF operates in a type of hierarchy.

The largest entity within the hierarchy is the autonomous system (AS), which is a collection of networks under a common administration that share a common routing strategy. OSPF is an intra-AS (interior gateway) routing protocol, although it is capable of receiving routes from and sending routes to other ASs.

You can divide an AS into a number of areas, which are groups of contiguous networks and attached hosts. Routers with multiple interfaces can participate in multiple areas. These routers, called area border routers (ABRs), maintain separate databases for each area. Areas are a logical grouping of OSPF routers identified by an integer or dotted-decimal number.

Areas allow you to further organize your routers within in the AS. One or more areas are required within the AS. Areas are valuable in that they allow sub-networks to "hide" within the AS, thus minimizing the size of the routing tables on all routers. An area within the AS may not see the details of another area’s topology. AS areas are known by their area number or the router’s IP address.

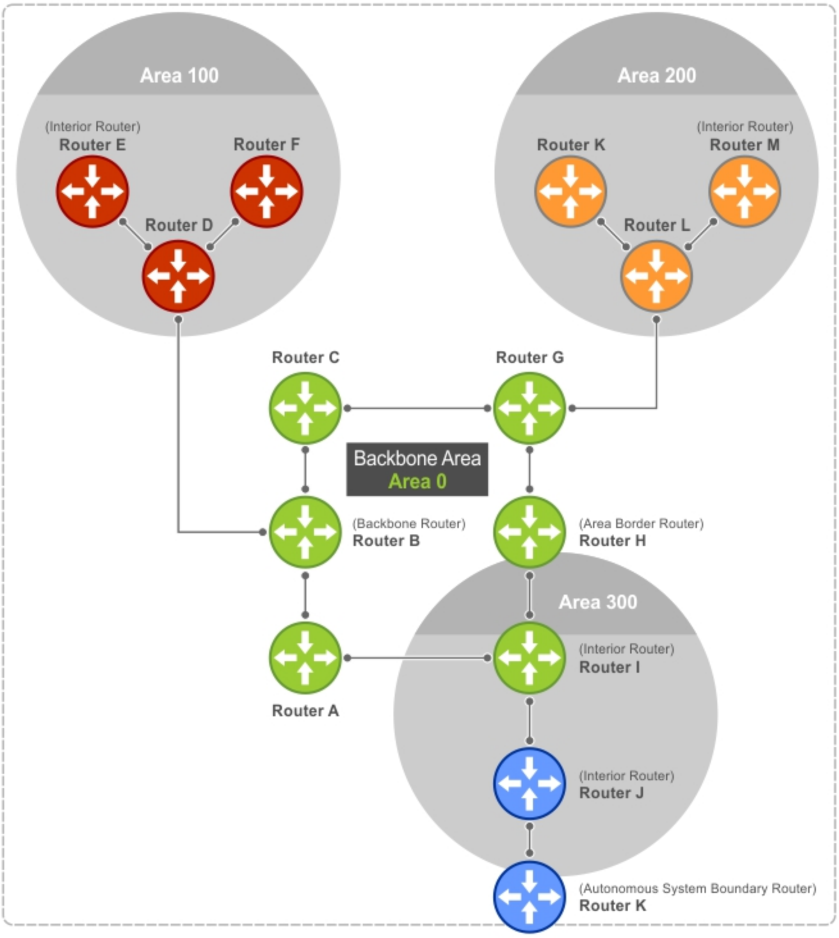

Area 100
(Interior Router)
Router E
Router F
Router D
Area 200
Router K
(Interior Router)
Router M
Router L
Router C
Router G
Backbone Area
Area 0
(Backbone Router)
Router B
(Area Border Router)
Router H
Area 300
(Interior Router)
Router I
Router A
(Interior Router)
Router J
(Autonomous System Boundary Router)
Router K

Figure 90. Autonomous System Areas

Area Types

The backbone of the network is Area 0. It is also called Area 0.0.0.0 and is the core of any AS.

All other areas must connect to Area 0. Areas can be defined in such a way that the backbone is not contiguous. In this case, backbone connectivity must be restored through virtual links. Virtual links are configured between any backbone routers that share a link to a non-backbone area and function as if they were direct links.

An OSPF backbone is responsible for distributing routing information between areas. It consists of all area border routers, networks not wholly contained in any area, and their attached routers.

NOTE: If you configure two non-backbone areas, then you must enable the B bit in OSPF.

The backbone is the only area with a default area number. All other areas can have their Area ID assigned in the configuration.

In the previous example, Routers A, B, C, G, H, and I are the Backbone.

- A stub area (SA) does not receive external route information, except for the default route. These areas do receive information from inter-area (IA) routes.

  NOTE: Configure all routers within an assigned stub area as stubby, and not generate LSAs that do not apply. For example, a Type 5 LSA is intended for external areas and the Stubby area routers may not generate external LSAs. A virtual link cannot traverse stubby areas.
- A not-so-stubby area (NSSA) can import AS external route information and send it to the backbone. It cannot receive external AS information from the backbone or other areas. However, a virtual link can traverse it.
- Totally stubby areas are referred to as no summary areas in the Dell Networking OS.

Networks and Neighbors

As a link-state protocol, OSPF sends routing information to other OSPF routers concerning the state of the links between them. The state (up or down) of those links is important.

Routers that share a link become neighbors on that segment. OSPF uses the Hello protocol as a neighbor discovery and keep alive mechanism. After two routers are neighbors, they may proceed to exchange and synchronize their databases, which creates an adjacency.

Router Types

Router types are attributes of the OSPF process.

A given physical router may be a part of one or more OSPF processes. For example, a router connected to more than one area, receiving routing from a border gateway protocol (BGP) process connected to another AS acts as both an area border router and an autonomous system router.

Each router has a unique ID, written in decimal format (A.B.C.D). You do not have to associate the router ID with a valid IP address. However, to make troubleshooting easier, Dell Networking recommends that the router ID and the router’s IP address reflect each other.

The following example shows different router designations.

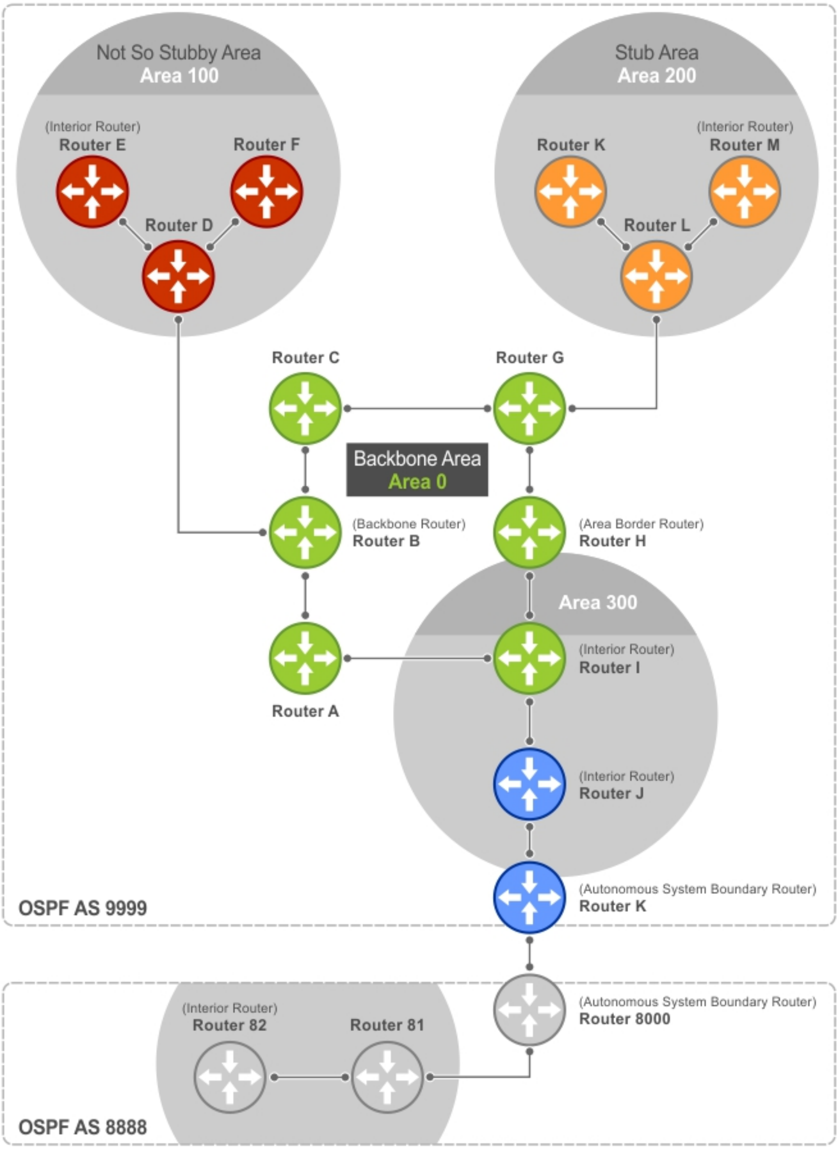

Not So Stubby Area
Area 100
(Interior Router)
Router E
Router F
Router D
Stub Area
Area 200
Router K
(Interior Router)
Router M
Router L
Router C
Router G
Backbone Area
Area 0
(Backbone Router)
Router B
(Area Border Router)
Router H
Area 300
(Interior Router)
Router I
Router A
(Interior Router)
Router J
(Autonomous System Boundary Router)
Router K
OSPF AS 9999
(Autonomous System Boundary Router)
Router 8000
(Interior Router)
Router 82
Router 81
OSPF AS 8888

Figure 91. OSPF Routing Examples

Backbone Router (BR)

A backbone router (BR) is part of the OSPF Backbone, Area 0.

This includes all ABRs. It can also include any routers that connect only to the backbone and another ABR, but are only part of Area 0, such as Router I in the previous example.

Area Border Router (ABR)

Within an AS, an area border router (ABR) connects one or more areas to the backbone.

The ABR keeps a copy of the link-state database for every area it connects to, so it may keep multiple copies of the link state database. An ABR takes information it has learned on one of its attached areas and can summarize it before sending it out on other areas it is connected to.

An ABR can connect to many areas in an AS, and is considered a member of each area it connects to.

Autonomous System Border Router (ASBR)

The autonomous system border area router (ASBR) connects to more than one AS and exchanges information with the routers in other ASs.

Generally, the ASBR connects to a non-interior gate protocol (IGP) such as BGP or uses static routes.

Internal Router (IR)

The internal router (IR) has adjacencies with ONLY routers in the same area, as Router E, M, and I shown in the previous example.

Designated and Backup Designated Routers

OSPF elects a designated router (DR) and a backup designated router (BDR). Among other things, the DR is responsible for generating LSAs for the entire multiaccess network.

Designated routers allow a reduction in network traffic and in the size of the topological database.

- The DR maintains a complete topology table of the network and sends the updates to the other routers via multicast. All routers in an area form a slave/master relationship with the DR. Every time a router sends an update, the router sends it to the DR and BDR. The DR sends the update out to all other routers in the area.
- The BDR is the router that takes over if the DR fails.

Each router exchanges information with the DR and BDR. The DR and BDR relay the information to the other routers. On broadcast network segments, the number of OSPF packets is further reduced by the DR and BDR sending such OSPF updates to a multicast IP address that all OSPF routers on the network segment are listening on.

These router designations are not the same ad the router IDs described earlier. The DRs and BDRs are configurable in Dell Networking OS. If you do not define DR or BDR in Dell Networking OS, the system assigns them. OSPF looks at the priority of the routers on the segment to determine which routers are the DR and BDR. The router with the highest priority is elected the DR. If there is a tie, the router with the higher router ID takes precedence. After the DR is elected, the BDR is elected the same way. A router with a router priority set to zero cannot become the DR or BDR.

Link-State Advertisements (LSAs)

A link-state advertisement (LSA) communicates the router's local routing topology to all other local routers in the same area.

The LSA types supported by Dell Networking are defined as follows:

- **Type 1: Router LSA** — The router lists links to other routers or networks in the same area. Type 1 LSAs are flooded across their own area only. The link-state ID of the Type 1 LSA is the originating router ID.
- **Type 2: Network LSA** — The DR in an area lists which routers are joined within the area. Type 2 LSAs are flooded across their own area only. The link-state ID of the Type 2 LSA is the IP interface address of the DR.
- **Type 3: Summary LSA (OSPFv2), Inter-Area-Prefix LSA (OSPFv3)** — An ABR takes information it has learned on one of its attached areas and can summarize it before sending it out on other areas it is connected to. The link-state ID of the Type 3 LSA is the destination network number.
- **Type 4: AS Border Router Summary LSA (OSPFv2), Inter-Area-Router LSA (OSPFv3)** — In some cases, Type 5 External LSAs are flooded to areas where the detailed next-hop information may not be

available. An ABR floods the information for the router (for example, the ASBR where the Type 5 advertisement originated. The link-state ID for Type 4 LSAs is the router ID of the described ASBR).

- **Type 5: LSA** — These LSAs contain information imported into OSPF from other routing processes. They are flooded to all areas, except stub areas. The link-state ID of the Type 5 LSA is the external network number.
- **Type 7: External LSA** — Routers in an NSSA do not receive external LSAs from ABRs, but are allowed to send external routing information for redistribution. They use Type 7 LSAs to tell the ABRs about these external routes, which the ABR then translates to Type 5 external LSAs and floods as normal to the rest of the OSPF network.
- **Type 8: Link LSA (OSPFv3)** — This LSA carries the IPv6 address information of the local links.
- **Type 9: Link Local LSA (OSPFv2), Intra-Area-Prefix LSA (OSPFv3)** — For OSPFv2, this is a link-local "opaque" LSA as defined by RFC2370. For OSPFv3, this LSA carries the IPv6 prefixes of the router and network links.
- **Type 11 - Grace LSA (OSPFv3)** — For OSPFv3 only, this LSA is a link-local “opaque” LSA sent by a restarting OSPFv3 router during a graceful restart.

For all LSA types, there are 20-byte LSA headers. One of the fields of the LSA header is the link-state ID.

Each router link is defined as one of four types: type 1, 2, 3, or 4. The LSA includes a link ID field that identifies, by the network number and mask, the object this link connects to.

Depending on the type, the link ID has different meanings.

- 1: point-to-point connection to another router/neighboring router.
- 2: connection to a transit network IP address of the DR.
- 3: connection to a stub network IP network/subnet number.
- 4: virtual link neighboring router ID.

LSA Throttling

LSA throttling provides configurable interval timers to improve OSPF convergence times.

The default OSPF static timers (5 seconds for transmission, 1 second for acceptance) ensures sufficient time for sending and resending LSAs and for system acceptance of arriving LSAs. However, some networks may require reduced intervals for LSA transmission and acceptance. Throttling timers allow for this improved convergence times.

The LSA throttling timers are configured in milliseconds, with the interval time increasing exponentially until a maximum time has been reached. If the maximum time is reached, the system, the system continues to transmit at the max-interval until twice the max-interval time has passed. At that point, the system reverts to the start-interval timer and the cycle begins again.

When you configure the LSA throttle timers, syslog messages appear, indicating the interval times, as shown below for the transmit timer (45000ms) and arrival timer (1000ms).

```
Mar 15 09:46:00: %STKUNIT0-M:CP %OSPF-4-LSA_BACKOFF: OSPF Process 10,Router lsa
id
2.2.2.2 router-id 2.2.2.2 is backed off to transmit after 45000ms

Mar 15 09:46:06: %STKUNIT0-M:CP %OSPF-4-LSA_BACKOFF: OSPF Process 10,Router lsa
id
3.3.3.3 rtrid 3.3.3.3 received before 1000ms time
```

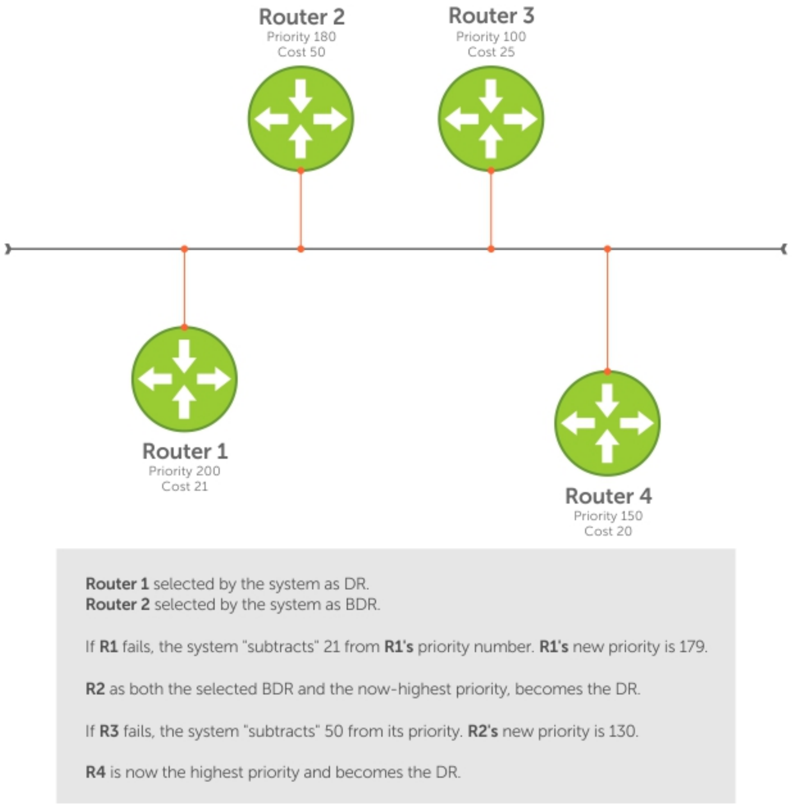
NOTE: The sequence numbers are reset when previously cleared routes that are waiting for the LSA throttle timer to expire are re-enabled.

Router Priority and Cost

Router priority and cost is the method the system uses to “rate” the routers.

For example, if not assigned, the system selects the router with the highest priority as the DR. The second highest priority is the BDR.

- Priority is a numbered rating 0 to 255. The higher the number, the higher the priority.
- Cost is a numbered rating 1 to 65535. The higher the number, the greater the cost. The cost assigned reflects the cost should the router fail. When a router fails and the cost is assessed, a new priority number results.



Router 1 selected by the system as DR.
Router 2 selected by the system as BDR.

If **R1** fails, the system "subtracts" 21 from **R1's** priority number. **R1's** new priority is 179.

R2 as both the selected BDR and the now-highest priority, becomes the DR.

If **R3** fails, the system "subtracts" 50 from its priority. **R2's** new priority is 130.

R4 is now the highest priority and becomes the DR.

Figure 92. Priority and Cost Examples

OSPF with Dell Networking OS

Dell Networking OS supports up to 10,000 OSPF routes for OSPFv2. Within that 10,000 routes, you can designate up to 8,000 routes as external and up to 2,000 as inter/intra area routes.

Dell Networking OS supports multiple OSPF processes (OSPF MP) on OSPFv2 only. The platform supports 32 OSPF processes simultaneously.

On OSPFv3, Dell Networking OS supports only one process at a time for all platforms.

OSPFv2 and OSPFv3 can coexist but you must configure them individually.

Dell Networking OS supports stub areas, totally stub (no summary) and not so stubby areas (NSSAs) and supports the following LSAs, as described earlier.

- Router (type 1)
- Network (type 2)
- Network Summary (type 3)
- AS Boundary (type 4)
- LSA(type 5)
- External LSA (type 7)
- Link LSA, OSPFv3 only (type 8)
- Opaque Link-Local (type 9)
- Grace LSA, OSPFv3 only (type 11)

Graceful Restart

When a router goes down without a graceful restart, there is a possibility for loss of access to parts of the network due to ongoing network topology changes. Additionally, LSA flooding and reconvergence can cause substantial delays. It is, therefore, desirable that the network maintains a stable topology if it is possible for data flow to continue uninterrupted.

OSPF graceful restart understands that in a modern router, the control plane and data plane functionality are separate, restarting the control plane functionality (such as the failover of the active RPM to the backup in a redundant configuration), does not necessarily have to interrupt the forwarding of data packets. This behavior is supported because the forwarding tables previously computed by an active RPM have been downloaded into the forwarding information base (FIB) on the line cards (the data plane) and are still resident. For packets that have existing FIB/CAM entries, forwarding between ingress and egress ports/VLANs, and so on, can continue uninterrupted while the control plane OSPF process comes back to full functionality and rebuilds its routing tables.

To notify its helper neighbors that the restart process is beginning, when a router is attempting to restart gracefully, it originates the following link-local Grace LSAs:

- An OSPFv2 router sends Type 9 LSAs.
- An OSPFv3 router sends Type 11 LSAs.

Type 9 and 11 LSAs include a grace period, which is the time period an OSPF router advertises to adjacent neighbor routers as the time to wait for it to return to full control plane functionality. During the grace period, neighbor OSPFv2 /v3 interfaces save the LSAs from the restarting OSPF interface. Helper neighbor routers continue to announce the restarting router as fully adjacent, as long as the network topology remains unchanged. When the restarting router completes its restart, it flushes the Type 9 and 11 LSAs, notifying its neighbors that the restart is complete. This notification happens before the grace period expires.

Dell Networking routers support the following OSPF graceful restart functionality:

- Restarting role in which an enabled router performs its own graceful restart.
- Helper role in which the router's graceful restart function is to help a restarting neighbor router in its graceful restarts.
- Helper-reject role in which OSPF does not participate in the graceful restart of a neighbor.

OSPFv2 supports *helper-only* and *restarting-only* roles. By default, both helper and restarting roles are enabled. OSPFv2 supports the helper-reject role globally on a router.

OSPFv3 supports *helper-only* and *restarting-only* roles. The helper-only role is enabled by default. To enable the restarting role in addition to the helper-only role, configure a grace period. Reconfigure OSPFv3 graceful restart to a restarting-only role when you enable the helper-reject role on an interface. OSPFv3 supports the helper-reject role on a per-interface basis.

Configuring helper-reject role on an OSPFv2 router or OSPFv3 interface enables the restarting-only role globally on the router or locally on the interface. In a helper-reject role, OSPF does not participate in the graceful restart of an adjacent OSPFv2/v3 router.

If multiple OSPF interfaces provide communication between two routers, after you configure helper-reject on one interface, all other interfaces between the two routers behave as if they are in the help-reject role.

- OSPFv2 and OSPFv3 support planned-only and/or unplanned-only restarts. The default is support for both planned and unplanned restarts.

  A planned restart occurs when you enter the `redundancy force-failover rpm` command to force the primary RPM to switch to the backup RPM. During a planned restart, OSPF sends out a Grace LSA before the system switches over to the backup RPM.

  An unplanned restart occurs when an unplanned event causes the active RPM to switch to the backup RPM, such as when an active process crashes, the active RPM is removed, or a power failure happens. During an unplanned restart, OSPF sends out a Grace LSA when the backup RPM comes online.

To display the configuration values for OSPF graceful restart, enter the `show run ospf` command for OSPFv2 and the `show run ospf` and `show ipv6 ospf database database-summary` commands for OSPFv3.

Fast Convergence (OSPFv2, IPv4 Only)

Fast convergence allows you to define the speeds at which LSAs are originated and accepted, and reduce OSPFv2 end-to-end convergence time.

Dell Networking OS allows you to accept and originate LSAa as soon as they are available to speed up route information propagation.

NOTE: The faster the convergence, the more frequent the route calculations and updates. This impacts CPU utilization and may impact adjacency stability in larger topologies.

Multi-Process OSPFv2 (IPv4 only)

Multi-process OSPF is supported on the S6000 platform with Dell Networking OS version 7.8.1.0 and later, and is supported on OSPFv2 with IPv4 only.

Multi-process OSPF allows multiple OSPFv2 processes on a single router. Multiple OSPFv2 processes allow for isolating routing domains, supporting multiple route policies and priorities in different domains, and creating smaller domains for easier management.

The S6000 supports up to 16 OSPFv2 processes.

Each OSPFv2 process has a unique process ID and must have an associated router ID. There must be an equal number of interfaces and must be in Layer-3 mode for the number of processes created. For example, if you create five OSPFv2 processes on a system, there must be at least five interfaces assigned in Layer 3 mode.

Each OSPFv2 process is independent. If one process loses adjacency, the other processes continue to function.

Processing SNMP and Sending SNMP Traps

Though there are may be several OSPFv2 processes, only one process can process simple network management protocol (SNMP) requests and send SNMP traps.

NOTE: SNMP get request corresponding to the OspfNbrOption feild in the OspfNbrTable returns a value of 66.

RFC-2328 Compliant OSPF Flooding

In OSPF, flooding is the most resource-consuming task. The flooding algorithm described in RFC 2328 requires that OSPF flood LSAs on all interfaces, as governed by LSA's flooding scope (refer to Section 13 of the RFC.)

When multiple direct links connect two routers, the RFC 2328 flooding algorithm generates significant redundant information across all links.

By default, Dell Networking OS implements an enhanced flooding procedure which dynamically and intelligently detects when to optimize flooding. Wherever possible, the OSPF task attempts to reduce flooding overhead by selectively flooding on a subset of the interfaces between two routers.

Enabling RFC-2328 Compliant OSPF Flooding

To enable OSPF flooding, use the following command.
When you enable this command, it configures Dell Networking OS to flood LSAs on all interfaces.

- Enable RFC 2328 flooding.

  ROUTER OSPF mode

  ```
  flood-2328
  ```

Example of Viewing the Debug Log for Flooding Behavior

To confirm RFC 2328 flooding behavior, use the `debug ip ospf packet` command. The following example shows no change in the updated packets (shown in bold).

ACKs 2 (shown in bold) is printed only for ACK packets.

```
00:10:41 : OSPF(1000:00):
Rcv. v:2 t:5(LSAck) l:64 Acks 2 rid:2.2.2.2
    aid:1500 chk:0xdbee aut:0 auk: keyid:0 from:Vl 1000
       LSType:Type-5 AS External id:160.1.1.0 adv:6.1.0.0 seq:0x8000000c
       LSType:Type-5 AS External id:160.1.2.0 adv:6.1.0.0 seq:0x8000000c
00:10:41 : OSPF(1000:00):
Rcv. v:2 t:5(LSAck) l:64 Acks 2 rid:2.2.2.2
    aid:1500 chk:0xdbee aut:0 auk: keyid:0 from:Vl 100
       LSType:Type-5 AS External id:160.1.1.0 adv:6.1.0.0 seq:0x8000000c
       LSType:Type-5 AS External id:160.1.2.0 adv:6.1.0.0 seq:0x8000000c
00:10:41 : OSPF(1000:00):
Rcv. v:2 t:4(LSUpd) l:100 rid:6.1.0.0
    aid:0 chk:0xccbd aut:0 auk: keyid:0 from:Gi 10/21
       Number of LSA:2
      LSType:Type-5 AS External(5) Age:1 Seq:0x8000000c id:170.1.1.0 Adv:6.1.0.0
        Netmask:255.255.255.0 fwd:0.0.0.0 E2, tos:0 metric:0
```

```
      LSType:Type-5 AS External(5) Age:1 Seq:0x8000000c id:170.1.2.0 Adv:6.1.0.0
        Netmask:255.255.255.0 fwd:0.0.0.0 E2, tos:0 metric:0
```

To confirm that you enabled RFC-2328–compliant OSPF flooding, use the `show ip ospf` command.

```
Dell#show ip ospf
Routing Process ospf 1 with ID 2.2.2.2
Supports only single TOS (TOS0) routes
It is an Autonomous System Boundary Router
It is Flooding according to RFC 2328
SPF schedule delay 5 secs, Hold time between two SPFs 10 secs
Number of area in this router is 1, normal 0 stub 0 nssa 1
--More--
```

OSPF ACK Packing

The OSPF ACK packing feature bundles multiple LS acknowledgements in a single packet, significantly reducing the number of ACK packets transmitted when the number of LSAs increases.

This feature also enhances network utilization and reduces the number of small ACK packets sent to a neighboring router. OSPF ACK packing is enabled by default and non-configurable.

Setting OSPF Adjacency with Cisco Routers

To establish an OSPF adjacency between Dell Networking and Cisco routers, the hello interval and dead interval must be the same on both routers.
In Dell Networking OS, the OSPF dead interval value is, by default, set to **40 seconds**, and is independent of the OSPF hello interval. Configuring a hello interval does not change the dead interval in Dell Networking OS. In contrast, the OSPF dead interval on a Cisco router is, by default, four times as long as the hello interval. Changing the hello interval on the Cisco router automatically changes the dead interval.

To ensure equal intervals between the routers, use the following command.

- Manually set the dead interval of the Dell Networking router to match the Cisco configuration.

  INTERFACE mode

  ```
  ip ospf dead-interval <x>
  ```

Examples of Setting and Viewing a Dead Interval

In the following example, the dead interval is set at 4x the hello interval (shown in bold).

```
Dell(conf)#int gi 2/2
Dell(conf-if-gi-2/2)#ip ospf hello-interval 20
Dell(conf-if-gi-2/2)#ip ospf dead-interval 80

Dell(conf-if-gi-2/2)#
```

In the following example, the dead interval is set at 4x the hello interval (shown in bold).

```
Dell (conf-if-gi-2/2)#ip ospf dead-interval 20
Dell (conf-if-gi-2/2)#do show ip os int gi1/3
GigabitEthernet 2/2 is up, line protocol is up
  Internet Address 20.0.0.1/24, Area 0
  Process ID 10, Router ID 1.1.1.2, Network Type BROADCAST, Cost: 1
  Transmit Delay is 1 sec, State DR, Priority 1
  Designated Router (ID) 1.1.1.2, Interface address 30.0.0.1
  Backup Designated Router (ID) 1.1.1.1, Interface address 30.0.0.2
  Timer intervals configured, Hello 20, Dead 80, Wait 20, Retransmit 5
  Hello due in 00:00:04
```

```
  Neighbor Count is 1, Adjacent neighbor count is 1
  Adjacent with neighbor 1.1.1.1 (Backup Designated Router)
Dell (conf-if-gi-2/2)#
```

Configuration Information

The interfaces must be in Layer-3 mode (assigned an IP address) and enabled so that they can send and receive traffic. The OSPF process must know about these interfaces.

To make the OSPF process aware of these interfaces, they must be assigned to OSPF areas.

You must configure OSPF GLOBALLY on the system in CONFIGURATION mode.

NOTE: Loop back routes are not installed in the Route Table Manager (RTM) as non-active routes.

OSPF features and functions are assigned to each router using the `CONFIG-INTERFACE` commands for each interface.

NOTE: By default, OSPF is disabled.

Configuration Task List for OSPFv2 (OSPF for IPv4)

Open shortest path first version 2 (OSPF for IPv4) is supported on the S6000 platform.
The following configuration tasks include two mandatory tasks and several optional tasks:

- Enabling OSPFv2 (mandatory)
- Assigning a Router ID
- Enabling Multi-Process OSPF
- Assigning an OSPFv2 Area (mandatory)
- Enable OSPFv2 on Interfaces
- Configuring Stub Areas
- Enabling Passive Interfaces
- Enabling Fast-Convergence
- Changing OSPFv2 Parameters on Interfaces
- Enabling OSPFv2 Authentication
- Creating Filter Routes
- Applying Prefix Lists
- Redistributing Routes
- Troubleshooting OSPFv2

1. Configure a physical interface. Assign an IP address, physical or Loopback, to the interface to enable Layer 3 routing.
2. Enable OSPF globally. Assign network area and neighbors.
3. Add interfaces or configure other attributes.

For a complete list of the OSPF commands, refer to the *OSPF* section in the *Dell Networking OS Command Line Reference Guide* document.

Enabling OSPFv2

To enable Layer 3 routing, assign an IP address to an interface (physical or Loopback). By default, OSPF, similar to all routing protocols, is disabled.
You *must* configure at least one interface for Layer 3 before enabling OSPFv2 globally.

If implementing multi-process OSPF, create an equal number of Layer 3 enabled interfaces and OSPF process IDs. For example, if you create four OSPFv2 process IDs, you must have four interfaces with Layer 3 enabled.

1. Assign an IP address to an interface.
   CONFIG-INTERFACE mode

   ```
   ip address ip-address mask
   ```

   The format is A.B.C.D/M.

   If you are using a Loopback interface, refer to Loopback Interfaces.
2. Enable the interface.
   CONFIG-INTERFACE mode

   ```
   no shutdown
   ```
3. Return to CONFIGURATION mode to enable the OSPFv2 process globally.
   CONFIGURATION mode

   ```
   router ospf process-id [vrf {vrf name}]
   ```

   - *vrf name*: enter the keyword `VRF` and the instance name to tie the OSPF instance to the VRF. All network commands under this OSPF instance are later tied to the VRF instance.

   The range is from 0 to 65535.

   The OSPF process ID is the identifying number assigned to the OSPF process. The router ID is the IP address associated with the OSPF process.

   After the OSPF process and the VRF are tied together, the OSPF process ID cannot be used again in the system.

   If you try to enter an OSPF process ID, or if you try to enable more OSPF processes than available Layer 3 interfaces, prior to assigning an IP address to an interface and setting the `no shutdown` command, the following message displays:

   ```
   C300(conf)#router ospf 1
   % Error: No router ID available.
   ```

Assigning a Router ID

In CONFIGURATION ROUTER OSPF mode, assign the router ID.
The router ID is not required to be the router's IP address. However, Dell Networking recommends using the IP address as the router ID for easier management and troubleshooting. Optional *process-id* commands are also described.

- Assign the router ID for the OSPFv2 process.
  CONFIG-ROUTER-OSPF-id mode

  ```
  router-id ip address
  ```
- Disable OSPF.
  CONFIGURATION mode

  ```
  no router ospf process-id
  ```

- Reset the OSPFv2 process.

  EXEC Privilege mode

  ```
  clear ip ospf process-id
  ```
- View the current OSPFv2 status.

  EXEC mode

  ```
  show ip ospf process-id
  ```

Example of Viewing the Current OSPFv2 Status

```
Dell#show ip ospf 55555
Routing Process ospf 55555 with ID 10.10.10.10
Supports only single TOS (TOS0) routes
SPF schedule delay 5 secs, Hold time between two SPFs 10 secs
Number of area in this router is 0, normal 0 stub 0 nssa 0
Dell#
```

Enabling Multi-Process OSPF (OSPFv2, IPv4 Only)

Multi-process OSPF allows multiple OSPFv2 processes on a single router.

For more information, refer to Multi-Process OSPF (OSPFv2, IPv4 Only)

When configuring a single OSPF process, follow the same steps previously described. Repeat them as often as necessary for the desired number of processes. After the process is created, all other configurations apply as usual.

1. Assign an IP address to an interface.

   CONFIG-INTERFACE mode

   ```
   ip address ip-address mask
   ```

   Format: A.B.C.D/M.

   If you are using a Loopback interface, refer to Loopback Interfaces.
2. Enable the interface.

   CONFIG-INTERFACE mode

   ```
   no shutdown
   ```
3. Return to CONFIGURATION mode to enable the OSPFv2 process globally.

   CONFIGURATION mode

   ```
   router ospf process-id [vrf]
   ```

   The range is from 0 to 65535.

   After the OSPF process and the VRF are tied together, the OSPF process ID cannot be used again in the system.

If you try to enable more OSPF processes than available Layer 3 interfaces, the following message displays:

```
C300(conf)#router ospf 1
% Error: No router ID available.
```

Assigning an OSPFv2 Area

After you enable OSPFv2, assign the interface to an OSPF area. Set up OSPF areas and enable OSPFv2 on an interface with the `network` command.
You must have at least one AS area: Area 0. This is the backbone area. If your OSPF network contains more than one area, configure a backbone area (Area ID 0.0.0.0). Any area besides Area 0 can have any number ID assigned to it.

The OSPFv2 process evaluates the `network` commands in the order they are configured. Assign the network address that is most explicit first to include all subnets of that address. For example, if you assign the network address 10.0.0.0 /8, you cannot assign the network address 10.1.0.0 /16 because it is already included in the first network address.

When configuring the `network` command, configure a network address and mask that is a superset of the IP subnet configured on the Layer-3 interface for OSPFv2 to use.

You can assign the area in the following step by a number or with an IP interface address.

- Enable OSPFv2 on an interface and assign a network address range to a specific OSPF area.
  CONFIG-ROUTER-OSPF-id mode

  ```
  network ip-address mask area area-id
  ```

  The IP Address Format is A.B.C.D/M.

  The area ID range is from 0 to 65535 or A.B.C.D/M.

Enable OSPFv2 on Interfaces

Enable and configure OSPFv2 on each interface (configure for Layer 3 protocol), and not shutdown.

You can also assign OSPFv2 to a Loopback interface as a virtual interface.

OSPF functions and features, such as MD5 Authentication, Grace Period, Authentication Wait Time, are assigned on a per interface basis.

NOTE: If using features like MD5 Authentication, ensure all the neighboring routers are also configured for MD5.

In the example below, an IP address is assigned to an interface and an OSPFv2 area is defined that includes the IP address of a Layer 3 interface.

The first bold lines assign an IP address to a Layer 3 interface, and the`no shutdown` command ensures that the interface is UP.

The second bold line assigns the IP address of an interface to an area.

Example of Enabling OSPFv2 and Assigning an Area to an Interface

```
Dell#(conf)#int gi 4/44
Dell(conf-if-gi-4/44)#ip address 10.10.10.10/24
Dell(conf-if-gi-4/44)#no shutdown
Dell(conf-if-gi-4/44)#ex
```

```
Dell(conf)#router ospf 1
Dell(conf-router_ospf-1)#network 1.2.3.4/24 area 0
Dell(conf-router_ospf-1)#network 10.10.10.10/24 area 1
Dell(conf-router_ospf-1)#network 20.20.20.20/24 area 2
Dell(conf-router_ospf-1)#
Dell#
```

Dell Networking recommends using the interface IP addresses for the OSPFv2 router ID for easier management and troubleshooting.

To view the configuration, use the `show config` command in CONFIGURATION ROUTER OSPF mode.

OSPF, by default, sends hello packets out to all physical interfaces assigned an IP address that is a subset of a network on which OSPF is enabled.

To view currently active interfaces and the areas assigned to them, use the `show ip ospf interface` command.

Example of Viewing Active Interfaces and Assigned Areas

```
Dell>show ip ospf 1 interface

GigabitEthernet 12/17 is up, line protocol is up
  Internet Address 10.2.2.1/24, Area 0.0.0.0
  Process ID 1, Router ID 11.1.2.1, Network Type BROADCAST, Cost: 1
  Transmit Delay is 1 sec, State DR, Priority 1
  Designated Router (ID) 11.1.2.1, Interface address 10.2.2.1
  Backup Designated Router (ID) 0.0.0.0, Interface address 0.0.0.0
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 00:00:04
  Neighbor Count is 0, Adjacent neighbor count is 0

GigabitEthernet 12/21 is up, line protocol is up
  Internet Address 10.2.3.1/24, Area 0.0.0.0
  Process ID 1, Router ID 11.1.2.1, Network Type BROADCAST, Cost: 1
  Transmit Delay is 1 sec, State BDR, Priority 1
  Designated Router (ID) 13.1.1.1, Interface address 10.2.3.2
  Backup Designated Router (ID) 11.1.2.1, Interface address 10.2.3.1
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 00:00:05
  Neighbor Count is 1, Adjacent neighbor count is 1
    Adjacent with neighbor 13.1.1.1 (Designated Router)
Dell>
```

Loopback interfaces also help the OSPF process. OSPF picks the highest interface address as the router-id and a Loopback interface address has a higher precedence than other interface addresses.

Example of Viewing OSPF Status on a Loopback Interface

```
Dell#show ip ospf 1 int

GigabitEthernet 13/23 is up, line protocol is up
  Internet Address 10.168.0.1/24, Area 0.0.0.1
  Process ID 1, Router ID 10.168.253.2, Network Type BROADCAST, Cost: 1
  Transmit Delay is 1 sec, State DROTHER, Priority 1
  Designated Router (ID) 10.168.253.5, Interface address 10.168.0.4
  Backup Designated Router (ID) 192.168.253.3, Interface address 10.168.0.2
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 00:00:08
  Neighbor Count is 3, Adjacent neighbor count is 2
  Adjacent with neighbor 10.168.253.5 (Designated Router)
  Adjacent with neighbor 10.168.253.3 (Backup Designated Router)
```

```
Loopback 0 is up, line protocol is up
  Internet Address 10.168.253.2/32, Area 0.0.0.1
  Process ID 1, Router ID 10.168.253.2, Network Type LOOPBACK, Cost: 1
Loopback interface is treated as a stub Host.
Dell#
```

Configuring Stub Areas

OSPF supports different types of LSAs to help reduce the amount of router processing within the areas. Type 5 LSAs are not flooded into stub areas; the ABR advertises a default route into the stub area to which it is attached. Stub area routers use the default route to reach external destinations.

To ensure connectivity in your OSPFv2 network, never configure the backbone area as a stub area.

To configure a stub area, use the following commands.

1. Review all areas after they were configured to determine which areas are NOT receiving type 5 LSAs.

   EXEC Privilege mode

   ```
   show ip ospf process-id [vrf] database database-summary
   ```
2. Enter CONFIGURATION mode.

   EXEC Privilege mode

   ```
   configure
   ```
3. Enter ROUTER OSPF mode.

   CONFIGURATION mode

   ```
   router ospf process-id [vrf]
   ```

   Process ID is the ID assigned when configuring OSPFv2 globally.
4. Configure the area as a stub area.

   CONFIG-ROUTER-OSPF-id mode

   ```
   area area-id stub [no-summary]
   ```

   Use the keywords `no-summary` to prevent transmission into the area of summary ASBR LSAs.

   Area ID is the number or IP address assigned when creating the area.

Example of the `show ip ospf database database-summary` Command

To view which LSAs are transmitted, use the `show ip ospf database` *`process-id`* `database-summary` command in EXEC Privilege mode.

```
Dell#show ip ospf 34 database database-summary

      OSPF Router with ID (10.1.2.100) (Process ID 34)

Area     ID Router Network S-Net S-ASBR Type-7 Subtotal
2.2.2.2  1         0       0     0      0      1
3.3.3.3  1         0       0     0      0      1
Dell#
```

To view information on areas, use the `show ip ospf` *`process-id`* command in EXEC Privilege mode.

Enabling Passive Interfaces

A passive interface is one that does not send or receive routing information.
Enabling passive interface suppresses routing updates on an interface. Although the passive interface does not send or receive routing updates, the network on that interface is still included in OSPF updates sent via other interfaces.

To suppress the interface's participation on an OSPF interface, use the following command. This command stops the router from sending updates on that interface.

- Specify whether all or some of the interfaces are passive.

  CONFIG-ROUTEROSPF- id mode

  ```
  passive-interface {default | interface}
  ```

  The default is enabled passive interfaces on ALL interfaces in the OSPF process.

  Entering the physical interface type, slot, and number enables passive interface on only the identified interface.

  - For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information (for example, `passive-interface gi 2/1`).
  - For a port channel, enter the keywords `port-channel` then a number from 1 to 255 for TeraScale and ExaScale.
  - For a SONET interface, enter the keyword `sonet` then the slot/port information (for example, `passive-interface so 2/2`).
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information (for example, `passive-interface ten 2/3`).
  - For a VLAN, enter the keyword `vlan` then a number from 1 to 4094 (for example, `passive-interface vlan 2222`).

  The keyword `default` sets all interfaces on this OSPF process as passive.

  To remove the passive interface from select interfaces, use the `no passive-interface` *`interface`* command while `passive interface default` is configured.

  To enable both receiving and sending routing updates, use the `no passive-interface` *`interface`* command.

Example of Viewing Passive Interfaces

When you configure a passive interface, the `show ip ospf` *`process-id`* `interface` command adds the words `passive interface` to indicate that the hello packets are not transmitted on that interface (shown in bold).

```
Dell#show ip ospf 34 int

GigabitEthernet 0/0 is up, line protocol is down
  Internet Address 10.1.2.100/24, Area 1.1.1.1
  Process ID 34, Router ID 10.1.2.100, Network Type BROADCAST, Cost: 10
  Transmit Delay is 1 sec, State DOWN, Priority 1
  Designated Router (ID) 10.1.2.100, Interface address 0.0.0.0
  Backup Designated Router (ID) 0.0.0.0, Interface address 0.0.0.0
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 13:39:46
  Neighbor Count is 0, Adjacent neighbor count is 0
```

```
GigabitEthernet 0/1 is up, line protocol is down
  Internet Address 10.1.3.100/24, Area 2.2.2.2
  Process ID 34, Router ID 10.1.2.100, Network Type BROADCAST, Cost: 10
  Transmit Delay is 1 sec, State DR, Priority 1
  Designated Router (ID) 10.1.2.100, Interface address 10.1.3.100
  Backup Designated Router (ID) 0.0.0.0, Interface address 0.0.0.0
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    No Hellos (Passive interface)
  Neighbor Count is 0, Adjacent neighbor count is 0

Loopback 45 is up, line protocol is up
  Internet Address 10.1.1.23/24, Area 2.2.2.2
  Process ID 34, Router ID 10.1.2.100, Network Type LOOPBACK, Cost: 1
```

Enabling Fast-Convergence

The fast-convergence CLI sets the minimum origination and arrival LSA parameters to zero (0), allowing rapid route calculation.
When you disable fast-convergence, origination and arrival LSA parameters are set to 5 seconds and 1 second, respectively.

Setting the convergence parameter (from 1 to 4) indicates the actual convergence level. Each convergence setting adjusts the LSA parameters to zero, but the `fast-convergence` parameter setting allows for even finer tuning of the convergence speed. The higher the number, the faster the convergence.

To enable or disable fast-convergence, use the following command.

- Enable OSPF fast-convergence and specify the convergence level.

  CONFIG-ROUTEROSPF- id mode

  ```
  fast-convergence {number}
  ```

  The parameter range is from 1 to 4.

  The higher the number, the faster the convergence.

  When disabled, the parameter is set at 0.

  NOTE: A higher convergence level can result in occasional loss of OSPF adjacency. Generally, convergence level 1 meets most convergence requirements. Only select higher convergence levels following consultation with Dell Technical Support.

Examples of the `fast-converge` Command

In the examples below, `Convergence Level` shows the fast-converge parameter setting and `Min LSA origination` shows the LSA parameters (shown in bold).

```
Dell(conf-router_ospf-1)#fast-converge 2
Dell(conf-router_ospf-1)#ex
Dell(conf)#ex
Dell#show ip ospf 1
Routing Process ospf 1 with ID 192.168.67.2
Supports only single TOS (TOS0) routes
SPF schedule delay 5 secs, Hold time between two SPFs 10 secs
Convergence Level 2
Min LSA origination 0 secs, Min LSA arrival 0 secs
Number of area in this router is 0, normal 0 stub 0 nssa 0
Dell#
```

The following examples shows how to disable fast-convergence.

```
Dell#(conf-router_ospf-1)#no fast-converge
Dell#(conf-router_ospf-1)#ex
Dell#(conf)#ex
Dell##show ip ospf 1
Routing Process ospf 1 with ID 192.168.67.2
Supports only single TOS (TOS0) routes
SPF schedule delay 5 secs, Hold time between two SPFs 10 secs
Convergence Level 0
Min LSA origination 5 secs, Min LSA arrival 1 secs
Number of area in this router is 0, normal 0 stub 0 nssa 0
Dell#
```

Changing OSPFv2 Parameters on Interfaces

In Dell Networking OS, you can modify the OSPF settings on the interfaces.
Some interface parameter values must be consistent across all interfaces to avoid routing errors. For example, set the same time interval for the hello packets on all routers in the OSPF network to prevent misconfiguration of OSPF neighbors.

To change OSPFv2 parameters on the interfaces, use any or all of the following commands.

- Change the cost associated with OSPF traffic on the interface.

  CONFIG-INTERFACE mode

  ```
  ip ospf cost
  ```

  - *`cost`*: The range is from 1 to 65535 (the default depends on the interface speed).
- Change the time interval the router waits before declaring a neighbor dead.

  CONFIG-INTERFACE mode

  ```
  ip ospf dead-interval seconds
  ```

  - *`seconds`*: the range is from 1 to 65535 (the default is **40 seconds**).

  The dead interval must be four times the hello interval.

  The dead interval must be the same on all routers in the OSPF network.
- Change the time interval between hello-packet transmission.

  CONFIG-INTERFACE mode

  ```
  ip ospf hello-interval seconds
  ```

  - *`seconds`*: the range is from 1 to 65535 (the default is **10 seconds**).

  The hello interval must be the same on all routers in the OSPF network.
- Use the MD5 algorithm to produce a message digest or key, which is sent instead of the key.

  CONFIG-INTERFACE mode

  ```
  ip ospf message-digest-key keyid md5 key
  ```

  - *`keyid`*: the range is from 1 to 255.
  - *`Key`*: a character string.

NOTE: Be sure to write down or otherwise record the key. You cannot learn the key after it is configured. You must be careful when changing this key.

NOTE: You can configure a maximum of six digest keys on an interface. Of the available six digest keys, the switches select the MD5 key that is common. The remaining MD5 keys are unused.

- Change the priority of the interface, which is used to determine the Designated Router for the OSPF broadcast network.

  CONFIG-INTERFACE mode

  ```
  ip ospf priority number
  ```

  - `number`: the range is from 0 to 255 (the default is **1**).
- Change the retransmission interval between LSAs.

  CONFIG-INTERFACE mode

  ```
  ip ospf retransmit-interval seconds
  ```

  - `seconds`: the range is from 1 to 65535 (the default is **5 seconds**).

  The retransmit interval must be the same on all routers in the OSPF network.
- Change the wait period between link state update packets sent out the interface.

  CONFIG-INTERFACE mode

  ```
  ip ospf transmit-delay seconds
  ```

  - `seconds`: the range is from 1 to 65535 (the default is **1 second**).

  The transmit delay must be the same on all routers in the OSPF network.

Example of Changing and Verifying the `cost` Parameter and Viewing Interface Status

To view interface configurations, use the `show config` command in CONFIGURATION INTERFACE mode.

To view interface status in the OSPF process, use the `show ip ospf interface` command in EXEC mode.

The bold lines in the example show the change on the interface. The change is reflected in the OSPF configuration.

```
Dell(conf-if)#ip ospf cost 45
Dell(conf-if)#show config
!
interface GigabitEthernet 0/0
  ip address 10.1.2.100 255.255.255.0
  no shutdown
  ip ospf cost 45
Dell(conf-if)#end

Dell#show ip ospf 34 interface
  GigabitEthernet 0/0 is up, line protocol is up
  Internet Address 10.1.2.100/24, Area 2.2.2.2
  Process ID 34, Router ID 10.1.2.100, Network Type BROADCAST, Cost: 45
  Transmit Delay is 1 sec, State DR, Priority 1
  Designated Router (ID) 10.1.2.100, Interface address 10.1.2.100
  Backup Designated Router (ID) 10.1.2.100, Interface address 0.0.0.0
```

```
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 00:00:06
  Neighbor Count is 0, Adjacent neighbor count is 0
Dell#
```

Enabling OSPFv2 Authentication

To enable or change various OSPF authentication parameters, use the following commands.

- Set a clear text authentication scheme on the interface.
  CONFIG-INTERFACE mode

  `ip ospf authentication-key *key*`

  Configure a *`key`* that is a text string no longer than eight characters.

  All neighboring routers must share password to exchange OSPF information.
- Set the authentication change wait time in seconds between 0 and 300 for the interface.
  CONFIG-INTERFACE mode

  `ip ospf auth-change-wait-time *seconds*`

  This setting is the amount of time OSPF has available to change its interface authentication type.

  When you configure the `auth-change-wait-time`, OSPF sends out only the old authentication scheme until the wait timer expires. After the wait timer expires, OSPF sends only the new authentication scheme. However, the new authentication scheme does not take effect immediately after the authentication change wait timer expires; OSPF accepts both the old as well as new authentication schemes for a time period that is equal to two times the configured authentication change wait timer. After this time period, OSPF accepts only the new authentication scheme.

  This transmission stops when the period ends.

  The default is **0 seconds**.

Enabling OSPFv2 Graceful Restart

Graceful restart is enabled for the global OSPF process.
For more information, refer to Graceful Restart.

The Dell Networking implementation of OSPFv2 graceful restart enables you to specify:

- **grace period** — the length of time the graceful restart process can last before OSPF terminates it.
- **helper-reject neighbors** — the router ID of each restart router that does not receive assistance from the configured router.
- **mode** — the situation or situations that trigger a graceful restart.
- **role** — the role or roles the configured router can perform.

NOTE: By default, OSPFv2 graceful restart is disabled.

To enable and configure OSPFv2 graceful restart, use the following commands.

1. Enable OSPFv2 graceful-restart globally and set the grace period.
   CONFIG-ROUTEROSPF- id mode

```
graceful-restart grace-period seconds
```

The seconds range is from 40 and 3000.

This setting is the time that an OSPFv2 router's neighbors advertises it as fully adjacent, regardless of the synchronization state, during a graceful restart. OSPFv2 terminates this process when the grace period ends.

2. Enter the Router ID of the OSPFv2 helper router from which the router does not accept graceful restart assistance.

   CONFIG-ROUTEROSPF- id mode

   ```
   graceful-restart helper-reject router-id
   ```

   - **Planned-only** — the OSPFv2 router supports graceful-restart for planned restarts only. A planned restart is when you manually enter a fail-over command to force the primary RPM over to the secondary RPM. During a planned restart, OSPF sends out a Grace LSA before the system switches over to the secondary RPM. OSPF also is notified that a planned restart is happening.
   - **Unplanned-only** — the OSPFv2 router supports graceful-restart for only unplanned restarts. During an unplanned restart, OSPF sends out a Grace LSA after the secondary RPM comes online.

   By default, OSPFv2 supports both planned and unplanned restarts. Selecting one or the other mode restricts OSPFv2 to the single selected mode.

3. Configure the graceful restart role or roles that this OSPFv2 router performs.

   CONFIG-ROUTEROSPF- id mode

   ```
   graceful-restart role [helper-only | restart-only]
   ```

   Dell Networking OS supports the following options:

   - `Helper-only`: the OSPFv2 router supports graceful-restart only as a helper router.
   - `Restart-only`: the OSPFv2 router supports graceful-restart only during unplanned restarts.

   By default, OSPFv2 supports both restarting and helper roles. Selecting one or the other role restricts OSPFv2 to the single selected role.

To disable OSPFv2 graceful-restart after you have enabled it, use the `no graceful-restart grace-period` command in CONFIG-ROUTEROSPF- id mode. The command returns OSPF graceful-restart to its default state.

NOTE: The Helper mode is enabled by default on the device. To enable the restart mode also on the device, you must configure the grace period using the `graceful-restart grace-period` command. After you enable restart mode the router advertises the neighbor as fully adjacent during a restart.

For more information about OSPF graceful restart, refer to the *Dell Networking OS Command Line Reference Guide*.

Example of the `show run ospf` Command

When you configure a graceful restart on an OSPFv2 router, the `show run ospf` command displays information similar to the following.

```
Dell#show run ospf
!
router ospf 1
```

```
  graceful-restart grace-period 300
  graceful-restart role helper-only
  graceful-restart mode unplanned-only
  graceful-restart helper-reject 10.1.1.1
  graceful-restart helper-reject 20.1.1.1
  network 10.0.2.0/24 area 0
Dell#
```

Creating Filter Routes

To filter routes, use prefix lists. OSPF applies prefix lists to incoming or outgoing routes.
Incoming routes must meet the conditions of the prefix lists. If they do not, OSPF does not add the route to the routing table. Configure the prefix list in CONFIGURATION PREFIX LIST mode prior to assigning it to the OSPF process.

- Create a prefix list and assign it a unique name.

  CONFIGURATION mode

  ```
  ip prefix-list prefix-name
  ```

  You are in PREFIX LIST mode.
- Create a prefix list with a sequence number and a deny or permit action.

  CONFIG- PREFIX LIST mode

  ```
  seq sequence-number {deny |permit} ip-prefix [ge min-prefix-length] [le max-prefix-length]
  ```

  The optional parameters are:
  - `ge min-prefix-length`: is the minimum prefix length to match (from 0 to 32).
  - `le max-prefix-length`: is the maximum prefix length to match (from 0 to 32).

For configuration information about prefix lists, refer to Access Control Lists (ACLs).

Applying Prefix Lists

To apply prefix lists to incoming or outgoing OSPF routes, use the following commands.

- Apply a configured prefix list to incoming OSPF routes.

  CONFIG-ROUTEROSPF-id mode

  ```
  distribute-list prefix-list-name in [interface]
  ```
- Assign a configured prefix list to outgoing OSPF routes.

  CONFIG-ROUTEROSPF-id

  ```
  distribute-list prefix-list-name out [connected | isis | rip | static]
  ```

Redistributing Routes

You can add routes from other routing instances or protocols to the OSPF process.
With the `redistribute` command, you can include RIP, static, or directly connected routes in the OSPF process.

NOTE: Do not route iBGP routes to OSPF unless there are route-maps associated with the OSPF redistribution.

To redistribute routes, use the following command.

- Specify which routes are redistributed into OSPF process.

  CONFIG-ROUTEROSPF-id mode

  ```
  redistribute {bgp | connected | isis | rip | static} [metric metric-value |
  metric-type type-value] [route-map map-name] [tag tag-value]
  ```

  Configure the following required and optional parameters:

  - `bgp, connected, isis, rip, static`: enter one of the keywords to redistribute those routes.
  - `metric` *`metric-value`*: the range is from 0 to 4294967295.
  - `metric-type` *`metric-type`*: 1 for OSPF external route type 1. 2 for OSPF external route type 2.
  - `route-map` *`map-name`*: enter a name of a configured route map.
  - `tag` *`tag-value`*: the range is from 0 to 4294967295.

Example of Viewing OSPF Configuration after Redistributing Routes

To view the current OSPF configuration, use the `show running-config ospf` command in EXEC mode or the `show config` command in ROUTER OSPF mode.

```
Dell(conf-router_ospf)#show config
!
router ospf 34
  network 10.1.2.32 0.0.0.255 area 2.2.2.2
  network 10.1.3.24 0.0.0.255 area 3.3.3.3
  distribute-list dilling in
Dell(conf-router_ospf)#
```

Troubleshooting OSPFv2

Dell Networking OS has several tools to make troubleshooting easier.
Be sure to check the following, as these questions represent typical issues that interrupt an OSPFv2 process.

NOTE: The following is not a comprehensive list, just some examples of typical troubleshooting checks.

- Have you enabled OSPF globally?
- Is the OSPF process active on the interface?
- Are adjacencies established correctly?
- Are the interfaces configured for Layer 3 correctly?
- Is the router in the correct area type?
- Have the routes been included in the OSPF database?
- Have the OSPF routes been included in the routing table (not just the OSPF database)?

Some useful troubleshooting commands are:

- `show interfaces`
- `show protocols`
- `debug IP OSPF events and/or packets`
- `show neighbors`
- `show virtual links`
- `show routes`

To help troubleshoot OSPFv2, use the following commands.

- View the summary of all OSPF process IDs enables on the router.

  EXEC Privilege mode

  `show running-config ospf`
- View the summary information of the IP routes.

  EXEC Privilege mode

  `show ip route summary`
- View the summary information for the OSPF database.

  EXEC Privilege mode

  `show ip ospf database`
- View the configuration of OSPF neighbors connected to the local router.

  EXEC Privilege mode

  `show ip ospf neighbor`
- View the LSAs currently in the queue.

  EXEC Privilege mode

  `show ip ospf timers rate-limit`
- View debug messages.

  EXEC Privilege mode

  `debug ip ospf` *`process-id`* `[event | packet | spf | database-timers rate-limit]`

  To view debug messages for a specific OSPF process ID, use the `debug ip ospf` *`process-id`* command.

  If you do not enter a process ID, the command applies to the first OSPF process.

  To view debug messages for a specific operation, enter one of the optional keywords:

  - `event`: view OSPF event messages.
  - `packet`: view OSPF packet information.
  - `spf`: view SPF information.
  - `database-timers rate-limit`: view the LSAs currently in the queue.

Example of Viewing OSPF Configuration

```
Dell#show run ospf
!
router ospf 3
!
router ospf 4
  router-id 4.4.4.4
  network 4.4.4.0/28 area 1
!
router ospf 5
!
router ospf 6
!
router ospf 7
  mib-binding
!
router ospf 8
```

```
!
router ospf 90
  area 2 virtual-link 4.4.4.4
  area 2 virtual-link 90.90.90.90 retransmit-interval 300
!
ipv6 router ospf 999
  default-information originate always
  router-id 10.10.10.10
Dell#
```

Sample Configurations for OSPFv2

The following configurations are examples for enabling OSPFv2.

These examples are not comprehensive directions. They are intended to give you some guidance with typical configurations.

You can copy and paste from these examples to your CLI. To support your own IP addresses, interfaces, names, and so on, be sure that you make the necessary changes.

Basic OSPFv2 Router Topology

The following illustration is a sample basic OSPFv2 topology.

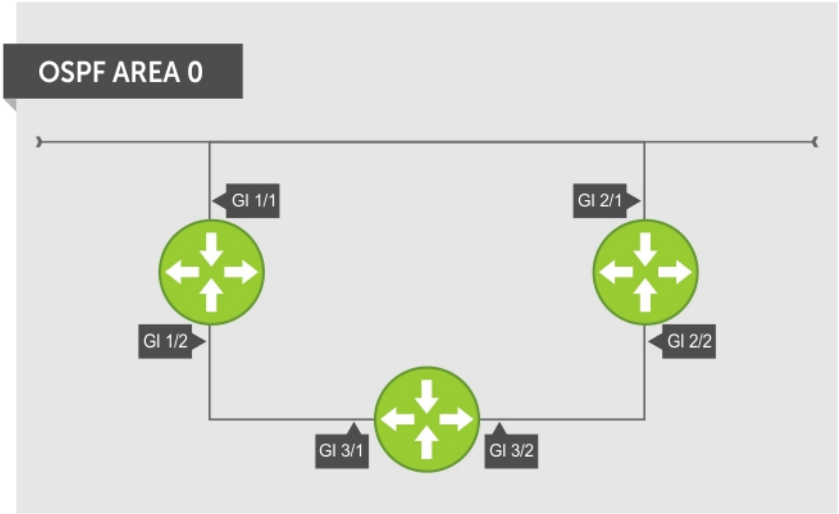


Figure 93. Basic Topology and CLI Commands for OSPFv2

OSPF Area 0 — Gl 1/1 and 1/2

```
router ospf 11111
  network 10.0.11.0/24 area 0
  network 10.0.12.0/24 area 0
  network 192.168.100.0/24 area 0
!
interface GigabitEthernet 1/1
  ip address 10.1.11.1/24
  no shutdown
!
interface GigabitEthernet 1/2
  ip address 10.2.12.2/24
  no shutdown
!
```

```
interface Loopback 10
  ip address 192.168.100.100/24
  no shutdown
```

OSPF Area 0 — Gl 3/1 and 3/2

```
router ospf 33333
  network 192.168.100.0/24 area 0
  network 10.0.13.0/24 area 0
  network 10.0.23.0/24 area 0
!
interface Loopback 30
  ip address 192.168.100.100/24
  no shutdown
!
interface GigabitEthernet 3/1
  ip address 10.1.13.3/24
  no shutdown
!
interface GigabitEthernet 3/2
  ip address 10.2.13.3/24
  no shutdown
```

OSPF Area 0 — Gl 2/1 and 2/2

```
router ospf 22222
  network 192.168.100.0/24 area 0
  network 10.2.21.0/24 area 0
  network 10.2.22.0/24 area 0
!
interface Loopback 20
  ip address 192.168.100.20/24
  no shutdown
!
interface GigabitEthernet 2/1
  ip address 10.2.21.2/24
  no shutdown
!
interface GigabitEthernet 2/2
  ip address 10.2.22.2/24
  no shutdown
```

Configuration Task List for OSPFv3 (OSPF for IPv6)

Open shortest path first version 3 (OSPF for IPv6) is supported on the platform.

The configuration options of OSPFv3 are the same as those options for OSPFv2, but you may configure OSPFv3 with differently labeled commands. Specify process IDs and areas and include interfaces and addresses in the process. Define areas as stub or totally stubby.

The interfaces must be in IPv6 Layer-3 mode (assigned an IPv6 IP address) and enabled so that they can send and receive traffic. The OSPF process must know about these interfaces. To make the OSPF process aware of these interfaces, assign them to OSPF areas.

The OSPFv3 `ipv6 ospf area` command enables OSPFv3 on the interface and places the interface in an area. With OSPFv2, two commands are required to accomplish the same tasks — the `router ospf` command to create the OSPF process, then the `network area` command to enable OSPF on an interface.

NOTE: The OSPFv2 `network area` command enables OSPF on multiple interfaces with the single command. Use the `OSPFv3 ipv6 ospf area` command on each interface that runs OSPFv3.

All IPv6 addresses on an interface are included in the OSPFv3 process that is created on the interface.

Enable OSPFv3 for IPv6 by specifying an OSPF process ID and an area in INTERFACE mode. If you have not created an OSPFv3 process, it is created automatically. All IPv6 addresses configured on the interface are included in the specified OSPF process.

NOTE: IPv6 and OSPFv3 do not support Multi-Process OSPF. You can only enable a single OSPFv3 process.

Enabling IPv6 Unicast Routing

To enable IPv6 unicast routing, use the following command.

- Enable IPv6 unicast routing globally.
  CONFIGURATION mode

  ```
  ipv6 unicast routing
  ```

Assigning IPv6 Addresses on an Interface

To assign IPv6 addresses to an interface, use the following commands.

1. Assign an IPv6 address to the interface.
   CONF-INT-type slot/port mode

   ```
   ipv6 address ipv6 address
   ```

   IPv6 addresses are normally written as eight groups of four hexadecimal digits; separate each group by a colon (:).

   The format is A:B:C::F/128.
2. Bring up the interface.
   CONF-INT-type slot/port mode

   ```
   no shutdown
   ```

Assigning Area ID on an Interface

To assign the OSPFv3 process to an interface, use the following command.
The `ipv6 ospf area` command enables OSPFv3 on an interface and places the interface in the specified area. Additionally, the command creates the OSPFv3 process with ID on the router. OSPFv2 requires two commands to accomplish the same tasks — the `router ospf` command to create the OSPF process, then the `network area` command to enable OSPFv2 on an interface.

NOTE: The OSPFv2 network area command enables OSPFv2 on multiple interfaces with the single command. Use the OSPFv3 `ipv6 ospf area` command on each interface that runs OSPFv3.

- Assign the OSPFv3 process and an OSPFv3 area to this interface.
  CONF-INT-type slot/port mode

```
ipv6 ospf process-id area area-id
```

  - *process-id*: the process ID number assigned.
  - *area-id*: the area ID for this interface.

Assigning OSPFv3 Process ID and Router ID Globally

To assign, disable, or reset OSPFv3 globally, use the following commands.

- Enable the OSPFv3 process globally and enter OSPFv3 mode.
  CONFIGURATION mode

  ```
  ipv6 router ospf {process ID}
  ```

  The range is from 0 to 65535.
- Assign the router ID for this OSPFv3 process.
  CONF-IPV6-ROUTER-OSPF mode

  ```
  router-id {number}
  ```

  - *number*: the IPv4 address.

  The format is A.B.C.D.

  NOTE: Enter the router-id for an OSPFv3 router as an IPv4 IP address.
- Disable OSPF.
  CONFIGURATION mode

  ```
  no ipv6 router ospf process-id
  ```
- Reset the OSPFv3 process.
  EXEC Privilege mode

  ```
  clear ipv6 ospf process
  ```

Enter an example that illustrates the current task (optional).
Enter the tasks the user should do after finishing this task (optional).

Configuring Stub Areas

To configure IPv6 stub areas, use the following command.

- Configure the area as a stub area.
  CONF-IPV6-ROUTER-OSPF mode

  ```
  area area-id stub [no-summary]
  ```

  - *no-summary*: use these keywords to prevent transmission in to the area of summary ASBR LSAs.
  - *Area ID*: a number or IP address assigned when creating the area. You can represent the area ID as a number from 0 to 65536 if you assign a dotted decimal format rather than an IP address.

Configuring Passive-Interface

To suppress the interface’s participation on an OSPFv3 interface, use the following command.
This command stops the router from sending updates on that interface.

- Specify whether some or all some of the interfaces are passive.

  CONF-IPV6-ROUTER-OSPF mode

  ```
  passive-interface {type slot/port}
  ```

  `Interface`: identifies the specific interface that is passive.

  - For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information (for example, `passive-interface gi 2/1`).
  - For a port channel, enter the keywords `port-channel` then a number from 1 to 255 for TeraScale and ExaScale (for example, `passive-interface po 100`)
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information (for example, `passive-interface ten 2/3`).
  - For a VLAN, enter the keyword `vlan` then a number from 1 to 4094 (for example, `passive-interface vlan 2222`).

To enable both receiving and sending routing updates, use the `no passive-interface interface` command.

To indicate that hello packets are not transmitted on that interface, when you configure a passive interface, the `show ipv6 ospf interface` command adds the words `passive interface`.

Redistributing Routes

You can add routes from other routing instances or protocols to the OSPFv3 process.

With the `redistribute` command, you can include RIP, static, or directly connected routes in the OSPF process. Route redistribution is also supported between OSPF Routing process IDs.

To add redistributing routes, use the following command.

- Specify which routes are redistributed into the OSPF process.

  CONF-IPV6-ROUTER-OSPF mode

  ```
  redistribute {bgp | connected | static} [metric metric-value | metric-type
  type-value] [route-map map-name] [tag tag-value]
  ```

  Configure the following required and optional parameters:

  - `bgp | connected | static`: enter one of the keywords to redistribute those routes.
  - `metric` *`metric-value`*: The range is from 0 to 4294967295.
  - `metric-type` *`metric-type`*: enter `1` for OSPFv3 external route type 1 OR `2` for OSPFv3 external route type 2.
  - `route-map` *`map-name`*: enter a name of a configured route map.
  - `tag` *`tag-value`*: The range is from 0 to 4294967295.

Configuring a Default Route

To generate a default external route into the OSPFv3 routing domain, configure Dell Networking OS.
To specify the information for the default route, use the following command.

- Specify the information for the default route.

  CONF-IPV6-ROUTER-OSPF mode

```
default-information originate [always [metric metric-value] [metric-type
type-value]] [route-map map-name]
```

Configure the following required and optional parameters:

- `always`: indicate that default route information is always advertised.
- `metric` *`metric-value`*: The range is from 0 to 4294967295.
- `metric-type` *`metric-type`*: enter `1` for OSPFv3 external route type 1 OR `2` for OSPFv3 external route type 2.
- `route-map` *`map-name`*: enter a name of a configured route map.

Enabling OSPFv3 Graceful Restart

Graceful restart for OSPFv3 is supported on the platform.
For more information about graceful restart, refer to Graceful Restart.

By default, OSPFv3 graceful restart is disabled and functions only in a helper role to help restarting neighbor routers in their graceful restarts when it receives a Grace LSA.

To enable OSPFv3 graceful restart, enter the `ipv6 router ospf process-id` command to enter OSPFv3 configuration mode. Then configure a grace period using the `graceful-restart grace-period` command. The grace period is the time that the OSPFv3 neighbors continue to advertise the restarting router as though it is fully adjacent. When you enable graceful restart (restarting role), an OSPFv3 restarting expects its OSPFv3 neighbors to help when it restarts by not advertising the broken link.

When you enable the helper-reject role on an interface using the `ipv6 ospf graceful-restart helper-reject` command, you reconfigure OSPFv3 graceful restart to function in a *restarting-only* role. OSPFv3 does not participate in the graceful restart of a neighbor.

NOTE: Enter the `ipv6 ospf graceful-restart helper-reject` command in Interface configuration mode.

- Enable OSPFv3 graceful restart globally by setting the grace period (in seconds).

  CONF-IPV6-ROUTER-OSPF mode

  ```
  graceful-restart grace-period seconds
  ```

  The valid values are from 40 to 1800 seconds.
- Configure an OSPFv3 interface to not act on the Grace LSAs that it receives from a restarting OSPFv3 neighbor.

  INTERFACE mode

  ```
  ipv6 ospf graceful-restart helper-reject
  ```
- Specify the operating mode and type of events that trigger a graceful restart.

  CONF-IPV6-ROUTER-OSPF mode

  ```
  graceful-restart mode [planned-only | unplanned-only]
  ```

  - `Planned-only`: the OSPFv3 router supports graceful restart only for planned restarts. A planned restart is when you manually enter a `redundancy force-failover rpm` command to force the primary RPM over to the secondary RPM. During a planned restart, OSPFv3 sends out a Grace LSA before the system switches over to the secondary RPM. OSPFv3 is notified that a planned restart is happening.

  - `Unplanned-only`: the OSPFv3 router supports graceful-restart only for unplanned restarts. During an unplanned restart, OSPFv3 sends out a Grace LSA once the secondary RPM comes online.

  The default is both planned and unplanned restarts trigger an OSPFv3 graceful restart. Selecting one or the other mode restricts OSPFv3 to the single selected mode.

- Disable OSPFv3 graceful-restart.

  CONF-IPV6-ROUTER-OSPF mode

  ```
  no graceful-restart grace-period
  ```

Displaying Graceful Restart

To display information on the use and configuration of OSPFv3 graceful restart, enter any of the following commands.

- Display the graceful-restart configuration for OSPFv2 and OSPFv3 (shown in the following example).

  EXEC Privilege mode

  ```
  show run ospf
  ```
- Display the Type-11 Grace LSAs sent and received on an OSPFv3 router (shown in the following example).

  EXEC Privilege mode

  ```
  show ipv6 ospf database grace-lsa
  ```
- Display the currently configured OSPFv3 parameters for graceful restart (shown in the following example).

  EXEC Privilege mode

  ```
  show ipv6 ospf database database-summary
  ```

Examples of the Graceful Restart `show` Commands

The following example shows the `show run ospf` command.

```
Dell#show run ospf
!
router ospf 1
  router-id 200.1.1.1
  log-adjacency-changes
  graceful-restart grace-period 180
  network 20.1.1.0/24 area 0
  network 30.1.1.0/24 area 0
!
ipv6 router ospf 1
  log-adjacency-changes
  graceful-restart grace-period 180
```

The following example shows the `show ipv6 ospf database database-summary` command.

```
Dell#show ipv6 ospf database database-summary
!
OSPFv3 Router with ID (200.1.1.1) (Process ID 1)

Process 1 database summary
Type                     Count/Status
Oper Status              1
Admin Status             1
Area Bdr Rtr Status      0
```

```
AS Bdr Rtr Status        1
AS Scope LSA Count       0
AS Scope LSA Cksum sum   0
Originate New LSAS       73
Rx New LSAS              114085
Ext LSA Count            0
Rte Max Eq Cost Paths    5
GR grace-period       180
GR mode             planned and unplanned

Area 0 database summary
Type                     Count/Status
Brd Rtr Count            2
AS Bdr Rtr Count         2
LSA count                12010
Summary LSAs             1
Rtr LSA Count            4
Net LSA Count            3
Inter Area Pfx LSA Count 12000
Inter Area Rtr LSA Count 0
Group Mem LSA Count      0
```

The following example shows the `show ipv6 ospf database grace-lsa` command.

```
Dell#show ipv6 ospf database grace-lsa
!
Type-11 Grace LSA (Area 0)

LS Age               : 10
Link State ID        : 6.16.192.66
Advertising Router   : 100.1.1.1
LS Seq Number        : 0x80000001
Checksum             : 0x1DF1
Length               : 36
Associated Interface : Gi 5/3
Restart Interval     : 180
Restart Reason       : Switch to Redundant Processor
```

OSPFv3 Authentication Using IPsec

OSPFv3 authentication using IP security (IPsec) is supported only on platform.

OSPFv3 uses IPsec to provide authentication for OSPFv3 packets. IPsec authentication ensures security in the transmission of OSPFv3 packets between IPsec-enabled routers.

IPsec is a set of protocols developed by the internet engineering task force (IETF) to support secure exchange of packets at the IP layer. IPsec supports two encryption modes: transport and tunnel.

- **Transport mode** — encrypts only the data portion (payload) of each packet, but leaves the header untouched.
- **Tunnel mode** — is more secure and encrypts both the header and payload. On the receiving side, an IPsec-compliant device decrypts each packet.

NOTE: Dell Networking OS supports only Transport Encryption mode in OSPFv3 authentication with IPsec.

With IPsec-based authentication, Crypto images are used to include the IPsec secure socket application programming interface (API) required for use with OSPFv3.

To ensure integrity, data origin authentication, detection and rejection of replays, and confidentiality of the packet, RFC 4302 and RFC 4303 propose using two security protocols — authentication header (AH) and encapsulating security payload (ESP). For OSPFv3, these two IPsec protocols provide interoperable, high-quality cryptographically-based security.

- **HA** — IPsec authentication header is used in packet authentication to verify that data is not altered during transmission and ensures that users are communicating with the intended individual or organization. Insert the authentication header after the IP header with a value of 51. AH provides integrity and validation of data origin by authenticating every OSPFv3 packet. For detailed information about the IP AH protocol, refer to *RFC 4302*.
- **ESP** — encapsulating security payload encapsulates data, enabling the protection of data that follows in the datagram. ESP provides authentication and confidentiality of every packet. The ESP extension header is designed to provide a combination of security services for both IPv4 and IPv6. Insert the ESP header after the IP header and before the next layer protocol header in Transport mode. It is possible to insert the ESP header between the next layer protocol header and encapsulated IP header in Tunnel mode. However, Tunnel mode is not supported in Dell Networking OS. For detailed information about the IP ESP protocol, refer to *RFC 4303*.

In OSPFv3 communication, IPsec provides security services between a pair of communicating hosts or security gateways using either AH or ESP. In an authentication policy on an interface or in an OSPF area, AH and ESP are used alone; in an encryption policy, AH and ESP may be used together. The difference between the two mechanisms is the extent of the coverage. ESP only protects IP header fields if they are encapsulated by ESP.

You decide the set of IPsec protocols that are employed for authentication and encryption and the ways in which they are employed. When you correctly implement and deploy IPsec, it does not adversely affect users or hosts. AH and ESP are designed to be cryptographic algorithm-independent.

OSPFv3 Authentication Using IPsec: Configuration Notes

OSPFv3 authentication using IPsec is implemented according to the specifications in RFC 4552.

- To use IPsec, configure an authentication (using AH) or encryption (using ESP) security policy on an interface or in an OSPFv3 area. Each security policy consists of a security policy index (SPI) and the key used to validate OSPFv3 packets. After IPsec is configured for OSPFv3, IPsec operation is invisible to the user.
  - You can only enable one security protocol (AH or ESP) at a time on an interface or for an area. Enable IPsec AH with the `ipv6 ospf authentication` command; enable IPsec ESP with the `ipv6 ospf encryption` command.
  - The security policy configured for an area is inherited by default on all interfaces in the area.
  - The security policy configured on an interface overrides any area-level configured security for the area to which the interface is assigned.
  - The configured authentication or encryption policy is applied to all OSPFv3 packets transmitted on the interface or in the area. The IPsec security associations (SAs) are the same on inbound and outbound traffic on an OSPFv3 interface.
  - There is no maximum AH or ESP header length because the headers have fields with variable lengths.
- Manual key configuration is supported in an authentication or encryption policy (dynamic key configuration using the internet key exchange [IKE] protocol is not supported).
- In an OSPFv3 authentication policy:
  - AH is used to authenticate OSPFv3 headers and certain fields in IPv6 headers and extension headers.
  - MD5 and SHA1 authentication types are supported; encrypted and unencrypted keys are supported.

- In an OSPFv3 encryption policy:
    - Both encryption and authentication are used.
    - IPsec security associations (SAs) are supported only in Transport mode (Tunnel mode is not supported).
    - ESP with null encryption is supported for authenticating only OSPFv3 protocol headers.
    - ESP with non-null encryption is supported for full confidentiality.
    - 3DES, DES, AES-CBC, and NULL encryption algorithms are supported; encrypted and unencrypted keys are supported.

NOTE: To encrypt all keys on a router, use the `service password-encryption` command in Global Configuration mode. However, this command does not provide a high level of network security. To enable key encryption in an IPsec security policy at an interface or area level, specify `7` for *`[key-encryption-type]`* when you enter the `ipv6 ospf authentication ipsec` or `ipv6 ospf encryption ipsec` command.

- To configure an IPsec security policy for authenticating or encrypting OSPFv3 packets on a physical, port-channel, or VLAN interface or OSPFv3 area, perform any of the following tasks:
    - Configuring IPsec Authentication on an Interface
    - Configuring IPsec Encryption on an Interface
    - Configuring IPsec Authentication for an OSPFv3 Area
    - Configuring IPsec Encryption for an OSPFv3 Area
    - Displaying OSPFv3 IPsec Security Policies

Configuring IPsec Authentication on an Interface

To configure, remove, or display IPsec authentication on an interface, use the following commands.

Prerequisite: Before you enable IPsec authentication on an OSPFv3 interface, first enable IPv6 unicast routing globally, configure an IPv6 address and enable OSPFv3 on the interface, and assign it to an area (refer to Configuration Task List for OSPFv3 (OSPF for IPv6)).

The SPI value must be unique to one IPsec security policy (authentication or encryption) on the router. Configure the same authentication policy (the same SPI and key) on each OSPFv3 interface in a link.

- Enable IPsec authentication for OSPFv3 packets on an IPv6-based interface.

  INTERFACE mode

  ```
  ipv6 ospf authentication {null | ipsec spi number {MD5 | SHA1} [key-encryption-type] key}
  ```

    - `null`: causes an authentication policy configured for the area to not be inherited on the interface.
    - `ipsec spi` *`number`*: the security policy index (SPI) value. The range is from 256 to 4294967295.
    - `MD5 | SHA1`: specifies the authentication type: Message Digest 5 (`MD5`) or Secure Hash Algorithm 1 (`SHA-1`).
    - *`key-encryption-type`*: (optional) specifies if the key is encrypted. The valid values are `0` (key is not encrypted) or `7` (key is encrypted).
    - `key`: specifies the text string used in authentication. All neighboring OSPFv3 routers must share key to exchange information. For MD5 authentication, the key must be 32 hex digits (non-encrypted) or 64 hex digits (encrypted). For SHA-1 authentication, the key must be 40 hex digits (non-encrypted) or 80 hex digits (encrypted).
- Remove an IPsec authentication policy from an interface.

```
no ipv6 ospf authentication ipsec spi number
```

- Remove null authentication on an interface to allow the interface to inherit the authentication policy configured for the OSPFv3 area.

  ```
  no ipv6 ospf authentication null
  ```

- Display the configuration of IPsec authentication policies on the router.

  ```
  show crypto ipsec policy
  ```

- Display the security associations set up for OSPFv3 interfaces in authentication policies.

  ```
  show crypto ipsec sa ipv6
  ```

Configuring IPsec Encryption on an Interface

To configure, remove, or display IPsec encryption on an interface, use the following commands.
Prerequisite: Before you enable IPsec encryption on an OSPFv3 interface, first enable IPv6 unicast routing globally, configure an IPv6 address and enable OSPFv3 on the interface, and assign it to an area (refer to Configuration Task List for OSPFv3 (OSPF for IPv6)).

NOTE: When you configure encryption using the `ipv6 ospf encryption ipsec` command, you enable both IPsec encryption and authentication. However, when you enable authentication on an interface using the `ipv6 ospf authentication ipsec` command, you do not enable encryption at the same time.

The SPI value must be unique to one IPsec security policy (authentication or encryption) on the router. Configure the same authentication policy (the same SPI and key) on each OSPFv3 interface in a link.

- Enable IPsec encryption for OSPFv3 packets on an IPv6-based interface.

  INTERFACE mode

  ```
  ipv6 ospf encryption {null | ipsec spi number esp encryption-algorithm [key-
  encryption-type] key authentication-algorithm [key-authentication-type] key}
  ```

  - `null`: causes an encryption policy configured for the area to not be inherited on the interface.
  - `ipsec spi` *`number`*: is the security policy index (SPI) value. The range is from 256 to 4294967295.
  - `esp` *`encryption-algorithm`*: specifies the encryption algorithm used with ESP. The valid values are `3DES`, `DES`, `AES-CBC`, and `NULL`. For `AES-CBC`, only the AES-128 and AES-192 ciphers are supported.
  - *`key`*: specifies the text string used in the encryption. All neighboring OSPFv3 routers must share the same key to decrypt information. Required lengths of a non-encrypted or encrypted key are: 3DES - 48 or 96 hex digits; DES - 16 or 32 hex digits; AES-CBC - 32 or 64 hex digits for AES-128 and 48 or 96 hex digits for AES-192.
  - *`key-encryption-type`*: (optional) specifies if the key is encrypted. The valid values are `0` (key is not encrypted) or `7` (key is encrypted).
  - *`authentication-algorithm`*: specifies the encryption authentication algorithm to use. The valid values are `MD5` or `SHA1`.
  - `key`: specifies the text string used in authentication. All neighboring OSPFv3 routers must share key to exchange information. For MD5 authentication, the key must be 32 hex digits (non-encrypted) or 64 hex digits (encrypted). For SHA-1 authentication, the key must be 40 hex digits (non-encrypted) or 80 hex digits (encrypted).
  - *`key-authentication-type`*: (optional) specifies if the authentication key is encrypted. The valid values are `0` or `7`.
- Remove an IPsec encryption policy from an interface.

  ```
  no ipv6 ospf encryption ipsec spi number
  ```

- Remove null encryption on an interface to allow the interface to inherit the encryption policy configured for the OSPFv3 area.

`no ipv6 ospf encryption null`

- Display the configuration of IPsec encryption policies on the router.

  `show crypto ipsec policy`
- Display the security associations set up for OSPFv3 interfaces in encryption policies.

  `show crypto ipsec sa ipv6`

Configuring IPSec Authentication for an OSPFv3 Area

To configure, remove, or display IPSec authentication for an OSPFv3 area, use the following commands. **Prerequisite**: Before you enable IPsec authentication on an OSPFv3 area, first enable OSPFv3 globally on the router (refer to Configuration Task List for OSPFv3 (OSPF for IPv6)).

The security policy index (SPI) value must be unique to one IPSec security policy (authentication or encryption) on the router. Configure the same authentication policy (the same SPI and key) on each interface in an OPSFv3 link.

If you have enabled IPSec encryption in an OSPFv3 area using the `area encryption` command, you cannot use the `area authentication` command in the area at the same time.

The configuration of IPSec authentication on an interface-level takes precedence over an area-level configuration. If you remove an interface configuration, an area authentication policy that has been configured is applied to the interface.

- Enable IPSec authentication for OSPFv3 packets in an area.

  CONF-IPV6-ROUTER-OSPF mode

  ```
  area-id authentication ipsec spi number {MD5 | SHA1} [key-encryption-type]
  key
  ```

  - `area area-id`: specifies the area for which OSPFv3 traffic is to be authenticated. For `area-id`, enter a number or an IPv6 prefix.
  - `spi number`: is the SPI value. The range is from 256 to 4294967295.
  - `MD5 | SHA1`: specifies the authentication type: message digest 5 (`MD5`) or Secure Hash Algorithm 1 (`SHA-1`).
  - `key-encryption-type`: (optional) specifies if the key is encrypted. The valid values are `0` (key is not encrypted) or `7` (key is encrypted).
  - `key`: specifies the text string used in authentication. All neighboring OSPFv3 routers must share key to exchange information. For MD5 authentication, the key must be 32 hex digits (non-encrypted) or 64 hex digits (encrypted). For SHA-1 authentication, the key must be 40 hex digits (non-encrypted) or 80 hex digits (encrypted).
- Remove an IPSec authentication policy from an OSPFv3 area.

  `no area area-id authentication ipsec spi number`
- Display the configuration of IPSec authentication policies on the router.

  `show crypto ipsec policy`

Configuring IPsec Encryption for an OSPFv3 Area

To configure, remove, or display IPsec encryption in an OSPFv3 area, use the following commands. **Prerequisite**: Before you enable IPsec encryption in an OSPFv3 area, first enable OSPFv3 globally on the router (refer to Configuration Task List for OSPFv3 (OSPF for IPv6)).

The SPI value must be unique to one IPsec security policy (authentication or encryption) on the router. Configure the same encryption policy (the same SPI and keys) on each interface in an OPSFv3 link.

NOTE: When you configure encryption using the `area encryption` command, you enable both IPsec encryption and authentication. However, when you enable authentication on an area using the `area authentication` command, you do not enable encryption at the same time.

If you have enabled IPsec authentication in an OSPFv3 area using the `area authentication` command, you cannot use the `area encryption` command in the area at the same time.

The configuration of IPsec encryption on an interface-level takes precedence over an area-level configuration. If you remove an interface configuration, an area encryption policy that has been configured is applied to the interface.

- Enable IPsec encryption for OSPFv3 packets in an area.

  CONF-IPV6-ROUTER-OSPF mode

  ```
  area area-id encryption ipsec spi number esp encryption-algorithm [key-
  encryption-type] key authentication-algorithm [key-authentication-type] key
  ```

  - `area area-id`: specifies the area for which OSPFv3 traffic is to be encrypted. For `area-id`, enter a number or an IPv6 prefix.
  - `spi number`: is the security policy index (SPI) value. The range is from 256 to 4294967295.
  - `esp encryption-algorithm`: specifies the encryption algorithm used with ESP. The valid values are `3DES`, `DES`, `AES-CBC`, and `NULL`. For `AES-CBC`, only the AES-128 and AES-192 ciphers are supported.
  - `key`: specifies the text string used in the encryption. All neighboring OSPFv3 routers must share the same key to decrypt information. The required lengths of a non-encrypted or encrypted key are: 3DES - 48 or 96 hex digits; DES - 16 or 32 hex digits; AES-CBC - 32 or 64 hex digits for AES-128 and 48 or 96 hex digits for AES-192.
  - `key-encryption-type`: (optional) specifies if the key is encrypted. Valid values: 0 (key is not encrypted) or 7 (key is encrypted).
  - `authentication-algorithm`: specifies the authentication algorithm to use for encryption. The valid values are `MD5` or `SHA1`.
  - `key`: specifies the text string used in authentication. All neighboring OSPFv3 routers must share key to exchange information. For `MD5` authentication, the key must be 32 hex digits (non-encrypted) or 64 hex digits (encrypted). For `SHA-1` authentication, the key must be 40 hex digits (non-encrypted) or 80 hex digits (encrypted).
  - `key-authentication-type`: (optional) specifies if the authentication key is encrypted. The valid values are `0` or `7`.

- Remove an IPsec encryption policy from an OSPFv3 area.

  ```
  no area area-id encryption ipsec spi number
  ```

- Display the configuration of IPsec encryption policies on the router.

  ```
  show crypto ipsec policy
  ```

Displaying OSPFv3 IPsec Security Policies

To display the configuration of IPsec authentication and encryption policies, use the following commands.

- Display the AH and ESP parameters configured in IPsec security policies, including the SPI number, key, and algorithms used.

  EXEC Privilege mode

  ```
  show crypto ipsec policy [name name]
  ```

  - `name`: displays configuration details about a specified policy.
- Display security associations set up for OSPFv3 links in IPsec authentication and encryption policies on the router.

  EXEC Privilege

  `show crypto ipsec sa ipv6 [interface *interface*]`

  To display information on the SAs used on a specific interface, enter `interface` *`interface`*, where interface is one of the following values:

  - For a 1-Gigabit Ethernet interface, enter `GigabitEthernet slot/port`.
  - For a Port Channel interface, enter `port-channel` *`number`*.
  - For a 10-Gigabit Ethernet interface, enter `TenGigabitEthernet slot/port`.
  - For a VLAN interface, enter `vlan` *`vlan-id`*. The valid VLAN IDs are from 1 to 4094.

Examples of the `show crypto ipsec` Commands

In the first example, the keys are not encrypted (shown in bold). In the second and third examples, the keys are encrypted (shown in bold).

The following example shows the `show crypto ipsec policy` command.

```
Dell#show crypto ipsec policy

Crypto IPSec client security policy data

Policy name             : OSPFv3-1-502
Policy refcount         : 1
Inbound ESP SPI         : 502 (0x1F6)
Outbound ESP SPI        : 502 (0x1F6)
Inbound ESP Auth Key    : 123456789a123456789b123456789c12
Outbound ESP Auth Key   : 123456789a123456789b123456789c12
Inbound ESP Cipher Key  : 123456789a123456789b123456789c123456789d12345678
Outbound ESP Cipher Key : 123456789a123456789b123456789c123456789d12345678
Transform set           : esp-3des esp-md5-hmac

Crypto IPSec client security policy data

Policy name             : OSPFv3-1-500
Policy refcount         : 2
Inbound AH SPI          : 500 (0x1F4)
Outbound AH SPI         : 500 (0x1F4)
Inbound AH Key          :
bbdd96e6eb4828e2e27bc3f9ff541e43faa759c9ef5706ba8ed8bb5efe91e97e
Outbound AH Key         :
bbdd96e6eb4828e2e27bc3f9ff541e43faa759c9ef5706ba8ed8bb5efe91e97e
Transform set           : ah-md5-hmac

Crypto IPSec client security policy data

Policy name             : OSPFv3-0-501
Policy refcount         : 1
Inbound ESP SPI         : 501 (0x1F5)
Outbound ESP SPI        : 501 (0x1F5)
Inbound ESP Auth Key    :
bbdd96e6eb4828e2e27bc3f9ff541e43faa759c9ef5706ba8ed8bb5efe91e97eb7c0c30808825fb5
Outbound ESP Auth Key   :
bbdd96e6eb4828e2e27bc3f9ff541e43faa759c9ef5706ba8ed8bb5efe91e97eb7c0c30808825fb5
Inbound ESP Cipher Key  :
bbdd96e6eb4828e2e27bc3f9ff541e43faa759c9ef5706ba10345a1039ba8f8a
Outbound ESP Cipher Key :
```

```
bbdd96e6eb4828e2e27bc3f9ff541e43faa759c9ef5706ba10345a1039ba8f8a
Transform set           : esp-128-aes esp-sha1-hmac
```

The following example shows the `show crypto ipsec sa ipv6` command.

```
Dell#show crypto ipsec sa ipv6

Interface: TenGigabitEthernet 0/0
  Link Local address: fe80::201:e8ff:fe40:4d10
  IPSecv6 policy name: OSPFv3-1-500

  inbound ah sas
   spi : 500 (0x1f4)
    transform : ah-md5-hmac
    in use settings : {Transport, }
    replay detection support : N
    STATUS : ACTIVE

  outbound ah sas
   spi : 500 (0x1f4)
    transform : ah-md5-hmac
    in use settings : {Transport, }
    replay detection support : N
    STATUS : ACTIVE

  inbound esp sas

  outbound esp sas

Interface: TenGigabitEthernet 0/1
  Link Local address: fe80::201:e8ff:fe40:4d11
  IPSecv6 policy name: OSPFv3-1-600

  inbound ah sas

  outbound ah sas

  inbound esp sas
   spi : 600 (0x258)
    transform : esp-des esp-sha1-hmac
    in use settings : {Transport, }
    replay detection support : N
    STATUS : ACTIVE

 outbound esp sas
  spi : 600 (0x258)
   transform : esp-des esp-sha1-hmac
   in use settings : {Transport, }
   replay detection support : N
   STATUS : ACTIVE
```

Troubleshooting OSPFv3

Dell Networking OS has several tools to make troubleshooting easier. Consider the following information as these are typical issues that interrupt the OSPFv3 process.

NOTE: The following troubleshooting section is not meant to be a comprehensive list, only examples of typical troubleshooting checks.

- Have you enabled OSPF globally?
- Is the OSPF process active on the interface?
- Are the adjacencies established correctly?

- Did you configure the interfaces for Layer 3 correctly?
- Is the router in the correct area type?
- Did you include the routes in the OSPF database?
- Did you include the OSPF routes in the routing table (not just the OSPF database)?

Some useful troubleshooting commands are:

- `show ipv6 interfaces`
- `show ipv6 protocols`
- `debug ipv6 ospf events and/or packets`
- `show ipv6 neighbors`
- `show virtual links`
- `show ipv6 routes`

Viewing Summary Information

To get general route, configuration, links status, and debug information, use the following commands.

- View the summary information of the IPv6 routes.

  EXEC Privilege mode

  ```
  show ipv6 route summary
  ```
- View the summary information for the OSPFv3 database.

  EXEC Privilege mode

  ```
  show ipv6 ospf database
  ```
- View the configuration of OSPFv3 neighbors.

  EXEC Privilege mode

  ```
  show ipv6 ospf neighbor
  ```
- View debug messages for all OSPFv3 interfaces.

  EXEC Privilege mode

  ```
  debug ipv6 ospf [event | packet] {type slot/port}
  ```

  - `event`: View OSPF event messages.
  - `packet`: View OSPF packets.
  - For a Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information (for example, `passive-interface gi 2/1`).
  - For a port channel, enter the keywords `port-channel` then a number from 1 to 255 for TeraScale and ExaScale.
  - For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information (for example, `passive-interface ten 2/3`).
  - For a VLAN, enter the keyword `vlan` then a number from 1 to 4094 (for example, `passive-interface vlan 2222`). FTOS supports 4094 VLANs.

32

Policy-based Routing (PBR)

Policy-based Routing (PBR) allows a switch to make routing decisions based on policies applied to an interface.

This chapter covers the following topics:

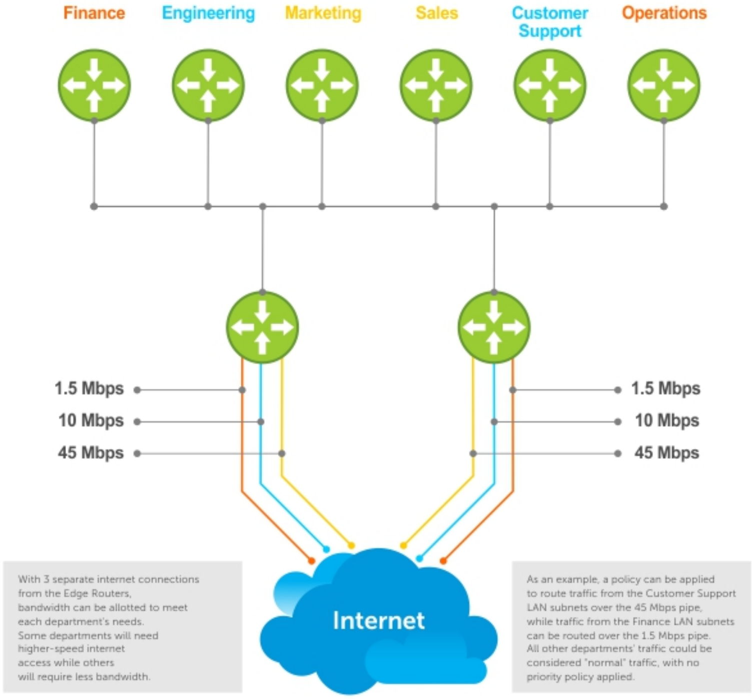
- Overview
- Implementing Policy-based Routing with Dell Networking OS
- Configuration Task List for Policy-based Routing
- Sample Configuration

Overview

When a router receives a packet it normally decides where to forward it based on the destination address in the packet, which is used to look up an entry in a routing table. However, in some cases, there may be a need to forward the packet based on other criteria: size, source, protocol type, destination, etc. For example, a network administrator might want to forward a packet that uses TCP across a different next-hop than packets using ICMP. In these situations, you can configure a switch route packets according to a policy applied to interfaces.

Rules for **PBR** can also be a combination of things:

When the packet comes from this source and wants to go to that destination then route it to this next-hop or onto that specific interface. This permits routing over different links or towards different networks even while the destination is the same but depending on where the packet originates.



With 3 separate internet connections from the Edge Routers, bandwidth can be allotted to meet each department's needs. Some departments will need higher-speed internet access while others will require less bandwidth.

As an example, a policy can be applied to route traffic from the Customer Support LAN subnets over the 45 Mbps pipe, while traffic from the Finance LAN subnets can be routed over the 1.5 Mbps pipe. All other departments' traffic could be considered "normal" traffic, with no priority policy applied.

To enable a PBR, you create a redirect list. Redirect lists are defined by rules, or routing policies. The following parameters can be defined in the routing policies or rules:

- IP address of the forwarding router (next-hop IP address)
- Protocol as defined in the header
- Source IP address and mask
- Destination IP address and mask
- Source port
- Destination port
- TCP Flags

Once a redirect-list is applied to an interface, all traffic passing through it is subjected to the rules defined in the redirect-list.

The traffic is forwarded based on the following:

- Next-hop addresses are verified. If the specified next hop is reachable, then the traffic is forwarded to the specified next-hop.
- If the specified next-hops are not reachable, then the normal routing table is used to forward the traffic.
- Dell Networking OS supports multiple next-hop entries in the redirect lists.
- Redirect-Lists are applied at Ingress.

Implementing Policy-based Routing with Dell Networking OS

- Non-contiguous bitmasks for PBR
- Hot-Lock PBR

Non-contiguous bitmasks for PBR

Non-contiguous bitmasks for PBR allows more granular and flexible control over routing policies. Network addresses that are in the middle of a subnet can be included or excluded. Specific bitmasks can be entered using the dotted decimal format.

Non-contiguous bitmask example

```
Dell#show ip redirect-list
IP redirect-list rcl0:
 Defined as:
  seq 5 permit ip 200.200.200.200 200.200.200.200 199.199.199.199 199.199.199.199
  seq 10 redirect 1.1.1.2 tcp 234.224.234.234 255.234.234.234 222.222.222.222/24
  seq 40 ack, Next-hop reachable(via Te 8/1), ARP resolved
 Applied interfaces:
  Te 8/0
```

Hot-Lock PBR

Ingress and egress Hot Lock PBR allow you to add or delete new rules into an existing policy (already written into CAM) without disruption to traffic flow. Existing entries in CAM are adjusted to accommodate the new entries. Hot Lock PBR is enabled by default.

Configuration Task List for Policy-based Routing

To enable the PBR:

- Create a Redirect List
- Create a Rule for a Redirect-list
- Apply a Redirect-list to an Interface using a Redirect-group

Create a Redirect List

Use the following command in **CONFIGURATION** mode:

Command Syntax	Command Mode	Purpose
ip redirect-list *redirect-list-name*	CONFIGURATION	Create a redirect list by entering the list name. Format: 16 characters

Delete the redirect list with the **no ip redirect-list** command.

The following example creates a redirect list by the name of “xyz.”

```
Dell(conf)#ip redirect-list ?
WORD   Redirect-list name (max 16 chars)
Dell(conf)#ip redirect-list xyz
```

Create a Rule for a Redirect-list

Use the following command in `CONFIGURATION REDIRECT-LIST` mode to set the rules for the redirect list. You can enter the command multiple times and create a sequence of redirect rules. Use the **seq** *nn* **redirect** version of the command to organize your rules**.**

Command Syntax	Command Mode	Purpose
seq *{number}* **redirect** {*ip-address*}{*ip-protocol-number* | *protocol-type* [*bit*]} {*source mask* | **any** | **host** *ip-address*} {*destination mask* | **any** | **host** *ip-address*}	CONF-REDIRECT-LIST	Configure a rule for the redirect list. *number* is the number in sequence to initiate this rule *ip-address* is the Forwarding router’s address FORMAT: A.B.C.D FORMAT: slot/port *ip-protocol-number* or *protocol-type* is the type of protocol to be redirected FORMAT: 0-255 for IP protocol number, or enter protocol type *source ip-address* or *any* or *host ip-address* is the Source’s IP address FORMAT: A.B.C.D/NN, or ANY or HOST IP address *destination ip-address* or *any* or *host ip-address* is the Destination’s IP address FORMAT: A.B.C.D/NN, or ANY or HOST IP address

Delete a rule with the **no redirect** command.

The redirect rule supports Non-contiguous bitmasks for PBR in the Destination router IP address

The below step shows a step-by-step example of how to create a rule for a redirect list by configuring:

- IP address of the next-hop router in the forwarding route
- IP protocol number
- Source address with mask information
- Destination address with mask information

Creating a Rule Example:

```
Dell(conf-redirect-list)#redirect ?
A.B.C.D                 Forwarding router's address
```

```
Dell(conf-redirect-list)#redirect 3.3.3.3 ?
<0-255>                 An IP protocol number
icmp                    Internet Control Message Protocol
ip                      Any Internet Protocol
tcp                     Transmission Control Protocol
udp                     User Datagram Protocol
Dell(conf-redirect-list)#redirect 3.3.3.3 ip ?
A.B.C.D                 Source address
any                     Any source host
host                    A single source host
Dell(conf-redirect-list)#redirect 3.3.3.3 ip 222.1.1.1 ?
Mask                    A.B.C.D or /nn Mask in dotted decimal or in slash
format
Dell(conf-redirect-list)#redirect 3.3.3.3 ip 222.1.1.1 /32 ?
A.B.C.D                 Destination address
any                     Any destination host
host                    A single destination host
Dell(conf-redirect-list)#redirect 3.3.3.3 ip 222.1.1.1 /32 77.1.1.1 ?
Mask                    A.B.C.D or /nn Mask in dotted decimal or in slash
format
Dell(conf-redirect-list)#redirect 3.3.3.3 ip 222.1.1.1 /32 77.1.1.1 /32 ?
Dell(conf-redirect-list)#redirect 3.3.3.3 ip 222.1.1.1 /32 77.1.1.1 /32
Dell(conf-redirect-list)#do show ip redirect-list

IP redirect-list xyz:
 Defined as:
  seq 5 redirect 3.3.3.3 ip host 222.1.1.1 host 77.1.1.1
  Applied interfaces:
  None
```

Multiple rules can be applied to a single redirect-list. The rules are applied in ascending order, starting with the rule that has the lowest sequence number in a redirect-list displays the correct method for applying multiple rules to one list.

Creating multiple rules for a redirect-list:

```
Dell(conf)#ip redirect-list test
Dell(conf-redirect-list)#seq 10 redirect 10.1.1.2 ip 20.1.1.0/24 any
Dell(conf-redirect-list)#seq 15 redirect 10.1.1.3 ip 20.1.1.0/25 any
Dell(conf-redirect-list)#seq 20 redirect 10.1.1.3 ip 20.1.1.128/24 any
Dell(conf-redirect-list)#show config
!
ip redirect-list test
 seq 10 redirect 10.1.1.2 ip 20.1.1.0/24 any
 seq 15 redirect 10.1.1.3 ip 20.1.1.0/25 any
 seq 20 redirect 10.1.1.3 ip 20.1.1.0/24 any
Dell(conf-redirect-list)#
```

NOTE: Starting in release 9.4(0.0), Dell Networking OS supports the use of multiple recursive routes with the same source-address and destination-address combination in a redirect policy on an router.

A recursive route is a route for which the immediate next-hop address is learned dynamically through a routing protocol and acquired through a route lookup in the routing table. The user can configure multiple recursive routes in a redirect list by entering multiple **seq redirect** statements with the same source and destination address and specify a different next-hop IP address. In this way, the recursive routes are used as different forwarding routes for dynamic failover. If the primary path goes down and the recursive route is removed from the routing table, the **seq redirect** statement is ignored and the next statement in the list with a different route is used.

PBR Exceptions (Permit)

Use the command **permit** to create an exception to a redirect list. Exceptions are used when a forwarding decision should be based on the routing table rather than a routing policy.

Dell Networking OS assigns the first available sequence number to a rule configured without a sequence number and inserts the rule into the PBR CAM region next to the existing entries. Since the order of rules is important, ensure that you configure any necessary sequence numbers.

The permit statement is never applied because the redirect list covers all source and destination IP addresses.

Ineffective PBR Exception due to Low Sequence Number

```
ip redirect-list rcl0
seq 5 redirect 2.2.2.2 ip any any
seq 10 permit ip host 3.3.3.3 any
```

To ensure that the permit statement or PBR exception is effective, use a lower sequence number, as shown below:

```
ip redirect-list rcl0
seq 10 permit ip host 3.3.3.3 any
seq 15 redirect 2.2.2.2 ip any any
```

Apply a Redirect-list to an Interface using a Redirect-group

IP redirect lists are supported on physical interfaces as well as VLAN and port-channel interfaces.

NOTE: When you apply a redirect-list on a port-channel, when traffic is redirected to the next hop and the destination port-channel is shut down, the traffic is dropped. However, on the S-Series, the traffic redirected to the destination port-channel is sometimes switched.

Use the following command in`INTERFACE` mode to apply a redirect list to an interface. Multiple redirect-lists can be applied to a redirect-group. It is also possible to create two or more redirect-groups on one interface for backup purposes.

Command Syntax	Command Mode	Purpose
ip redirect-group *redirect-list-name*	INTERFACE	Apply a redirect list (policy-based routing) to an interface. *redirect-list-name* is the name of a redirect list to apply to this interface. FORMAT: up to 16 characters

Delete the redirect list from this interface with the **[no] ip redirect-group** command.

In this example, the list "xyz" is applied to the tenGigabitEthernet 4/0 interface.

Applying a Redirect-list to an Interface Example:

```
Dell(conf-if-te-2/0)#ip redirect-group xyz
Dell(conf-if-te-2/0)#
```

Applying a Redirect-list to an Interface Example:

```
Dell(conf-if-te-1/0)#ip redirect-group test
Dell(conf-if-te-1/0)#ip redirect-group xyz
Dell(conf-if-te-1/0)#show config
!
interface TenGigabitEthernet 1/0
 no ip address
 ip redirect-group test
 ip redirect-group xyz
 shutdown
Dell(conf-if-te-1/0)#
```

In addition to supporting multiple redirect-lists in a redirect-group, multiple redirect-groups are supported on a single interface. Dell Networking OS has the capability to support multiple groups on an interface for backup purposes.

Show Redirect List Configuration

To view the configuration redirect list configuration, use the following command in EXEC mode:

Command Syntax	Command Mode	Purpose
show ip redirect-list *redirect-list-name*	EXEC	View the redirect list configuration and the associated interfaces.
show cam pbr **show cam-usage**	EXEC	View the redirect list entries programmed in the CAM.

List the redirect list configuration using the **show ip redirect-list redirect-list-name** command. The non-contiguous mask is displayed in dotted format (x.x.x.x). The contiguous mask is displayed in /x format. Some sample outputs are shown below:

```
Dell#show ip redirect-list xyz

IP redirect-list xyz:
 Defined as:
  seq 5 redirect 3.3.3.3 ip host 222.1.1.1 host 77.1.1.1
```

Use the **show ip redirect-list** (without the list name) to display all the redirect-lists configured on the device.

```
Dell#show ip redirect-list

IP redirect-list rcl0:
 Defined as:
  seq 5 permit ip 200.200.200.200 200.200.200.200 199.199.199.199 199.199.199.199
  seq 10 redirect 1.1.1.2 tcp 234.224.234.234 255.234.234.234
222.222.222.222/24 eq 40 ack, Next-hop reachable
(via Te 2/1), ARP resolved
 Applied interfaces:
  Te 2/0
```

NOTE: If, the redirect-list is applied to an interface, the output of **show ip redirect-list redirect-list-name** command displays reachability and ARP status for the specified next-hop.

Showing CAM PBR Configuration Example :

```
Dell(conf-if-te-2/1)#do show cam pbr stack-unit 0 port-set 0

TCP Flag: Bit 5 - URG, Bit 4 - ACK, Bit 3 - PSH, Bit 2 - RST, Bit 1 - SYN, Bit
0 - FIN

Cam   Port VlanID Proto Tcp   Src   Dst   SrcIp    DstIp  Next-hop
Egress
Index Flag  Port  Port  MAC   Port
--------------------------------------------------------------------------------
---------------------------------
06080 0 N/A    IP    0x0   0 0 200.200.200.200 200.200.200.200 199.199.199.199
199.199.199.199 N/A  NA
06081 0 N/A    TCP   0x10  0  40 234.234.234.234 255.234.234.234
222.222.222.222/24  00:00:00:00:00:09 8/1
```

Sample Configuration

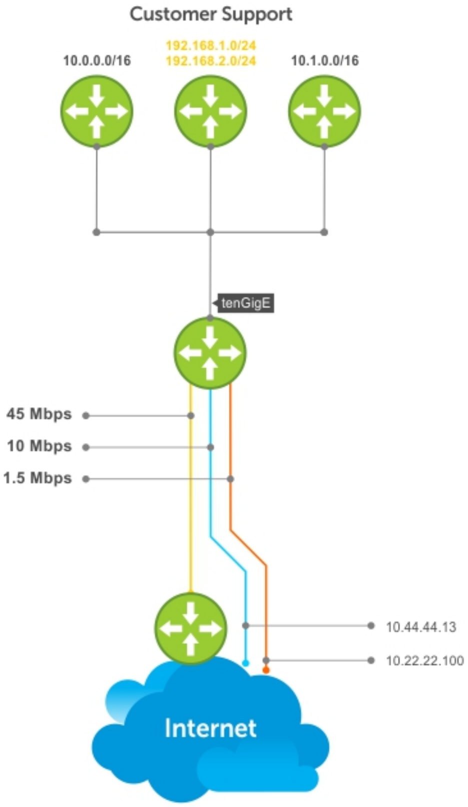
The following configuration is an example for setting up a PBR. These are not comprehensive directions. They are intended to give you a some guidance with typical configurations. You can copy and paste from these examples to your CLI. Be sure you make the necessary changes to support your own IP Addresses, Interfaces, Names, etc.

Graphic illustration of the configuration shown below:

The Redirect-List GOLD defined in this example, creates the following rules:

- description Route Gold traffic to the DS3.
- seq 5 redirect 10.99.99.254 ip 192.168.1.0/24 any " Redirect to next-hop router IP 10.99.99.254 any traffic originating in 192.168.1.0/24"
- seq 10 redirect 10.99.99.254 ip 192.168.2.0/24 any " Redirect to next-hop router IP 10.99.99.254 any traffic originating in 192.168.2.0/24"
- seq 15 permit ip any

PBR Sample Configuration examples are shown below:


Customer Support
10.0.0.0/16
192.168.1.0/24
192.168.2.0/24
10.1.0.0/16
tenGigE
45 Mbps
10 Mbps
1.5 Mbps
10.44.44.13
10.22.22.100
Internet

Create the Redirect-List GOLD

```
EDGE_ROUTER(conf-if-Te-2/23)#ip redirect-list GOLD
EDGE_ROUTER(conf-redirect-list)#description Route GOLD traffic to ISP_GOLD.
EDGE_ROUTER(conf-redirect-list)#direct 10.99.99.254 ip 192.168.1.0/24 any
EDGE_ROUTER(conf-redirect-list)#redirect 10.99.99.254 ip 192.168.2.0/24 any
EDGE_ROUTER(conf-redirect-list)# seq 15 permit ip any any
EDGE_ROUTER(conf-redirect-list)#show config
!
ip redirect-list GOLD
 description Route GOLD traffic to ISP_GOLD.
 seq 5 redirect 10.99.99.254 ip 192.168.1.0/24 any
 seq 10 redirect 10.99.99.254 ip 192.168.2.0/24 any
 seq 15 permit ip any any
```

Assign Redirect-List GOLD to Interface 2/11

```
EDGE_ROUTER(conf)#int Te 2/11
EDGE_ROUTER(conf-if-Te-2/11)#ip add 192.168.3.2/24
EDGE_ROUTER(conf-if-Te-2/11)#no shut
EDGE_ROUTER(conf-if-Te-2/11)#
EDGE_ROUTER(conf-if-Te-2/11)#ip redirect-group GOLD
EDGE_ROUTER(conf-if-Te-2/11)#no shut
EDGE_ROUTER(conf-if-Te-2/11)#end
EDGE_ROUTER(conf-redirect-list)#end


EDGE_ROUTER#
```

View Redirect-List GOLD

```
EDGE_ROUTER#show ip redirect-list

IP redirect-list GOLD:
 Defined as:
  seq 5 redirect 10.99.99.254 ip 192.168.1.0/24 any, Next-hop reachable (via Te
3/23), ARP resolved
  seq 10 redirect 10.99.99.254 ip 192.168.2.0/24 any, Next-hop reachable (via
Te 3/23), ARP resolved
  seq 15 permit ip any any
 Applied interfaces:
  Te 2/11
EDGE_ROUTER#
```

33

PIM Sparse-Mode (PIM-SM)

Protocol-independent multicast sparse-mode (PIM-SM) is supported on the S6000 platform.

PIM-SM is a multicast protocol that forwards multicast traffic to a subnet only after a request using a PIM Join message; this behavior is the opposite of PIM-Dense mode, which forwards multicast traffic to all subnets until a request to stop.

Implementation Information

Be aware of the following PIM-SM implementation information.

- The Dell Networking implementation of PIM-SM is based on IETF *Internet Draft draft-ietf-pim-sm-v2-new-05*.
- The SPT-Threshold is zero, which means that the last-hop designated router (DR) joins the shortest path tree (SPT) to the source after receiving the first multicast packet.
- Dell Networking OS reduces the number of control messages sent between multicast routers by bundling Join and Prune requests in the same message.
- Dell Networking OS supports PIM-SM on physical, virtual local area network (VLAN), and port-channel interfaces.
- Dell Networking OS supports 2000 IPv6 multicast forwarding entries, with up to 128 PIM-source-specific multicast (SSM) neighbors/interfaces.
- IPv6 Multicast is not supported on synchronous optical network technologies (SONET) interfaces.

Protocol Overview

PIM-SM initially uses unidirectional shared trees to forward multicast traffic; that is, all multicast traffic must flow only from the rendezvous point (RP) to the receivers.

After a receiver receives traffic from the RP, PM-SM switches to SPT to forward multicast traffic. Every multicast group has an RP and a unidirectional shared tree (group-specific shared tree).

Requesting Multicast Traffic

A host requesting multicast traffic for a particular group sends an Internet group management protocol (IGMP) Join message to its gateway router.

The gateway router is then responsible for joining the shared tree to the RP (RPT) so that the host can receive the requested traffic.

1. After receiving an IGMP Join message, the receiver gateway router (last-hop DR) creates a (*,G) entry in its multicast routing table for the requested group. The interface on which the join message was received becomes the outgoing interface associated with the (*,G) entry.
2. The last-hop DR sends a PIM Join message to the RP. All routers along the way, including the RP, create an (*,G) entry in their multicast routing table, and the interface on which the message was received becomes the outgoing interface associated with the (*,G) entry. This process constructs an RPT branch to the RP.

3. If a host on the same subnet as another multicast receiver sends an IGMP report for the same multicast group, the gateway takes no action. If a router between the host and the RP receives a PIM Join message for which it already has a (*,G) entry, the interface on which the message was received is added to the outgoing interface list associated with the (*,G) entry, and the message is not (and does not need to be) forwarded towards the RP.

Refuse Multicast Traffic

A host requesting to leave a multicast group sends an IGMP Leave message to the last-hop DR. If the host is the only remaining receiver for that group on the subnet, the last-hop DR is responsible for sending a PIM Prune message up the RPT to prune its branch to the RP.

1. After receiving an IGMP Leave message, the gateway removes the interface on which it is received from the outgoing interface list of the (*,G) entry. If the (*,G) entry has no remaining outgoing interfaces, multicast traffic for that group is no longer forwarded to that subnet.
2. If the (*,G) entry has no remaining outgoing interfaces, the last-hop DR sends a PIM Prune message to towards the RP. All routers along the way remove the interface on which the message was received from the outgoing interface list of the (*,G) entry. If on any router there is at least one outgoing interface listed for that (*,G) entry, the Prune message is not forwarded.

Send Multicast Traffic

With PIM-SM, all multicast traffic must initially originate from the RP. A source must unicast traffic to the RP so that the RP can learn about the source and create an SPT to it. Then the last-hop DR may create an SPT directly to the source.

1. The source gateway router (first-hop DR) receives the multicast packets and creates an (S,G) entry in its multicast routing table. The first-hop DR encapsulates the initial multicast packets in PIM Register packets and unicasts them to the RP.
2. The RP decapsulates the PIM Register packets and forwards them if there are any receivers for that group. The RP sends a PIM Join message towards the source. All routers between the RP and the source, including the RP, create an (S,G) entry and list the interface on which the message was received as an outgoing interface, thus recreating a SPT to the source.
3. After the RP starts receiving multicast traffic via the (S,G), it unicasts a Register-Stop message to the first-hop DR so that multicast packets are no longer encapsulated in PIM Register packets and unicast. After receiving the first multicast packet from a particular source, the last-hop DR sends a PIM Join message to the source to create an SPT to it.
4. There are two paths, then, between the receiver and the source, a direct SPT and an RPT. One router receives a multicast packet on two interfaces from the same source in this case; this router prunes the shared tree by sending a PIM Prune message to the RP that tells all routers between the source and the RP to remove the outgoing interface from the (*,G) entry, and tells the RP to prune its SPT to the source with a Prune message.

Dell Networking OS Behavior: When the router creates an SPT to the source, there are then two paths between the receiver and the source, the SPT and the RPT. Until the router can prune itself from the RPT, the receiver receives duplicate multicast packets which may cause disruption. Therefore, the router must prune itself from the RPT as soon as possible. Dell Networking OS optimizes the shared to shortest-path tree switchover latency by copying and forwarding the first (S,G) packet received on the SPT to the PIM task immediately upon arrival. The arrival of the (S,G) packet confirms for PIM that the SPT is created, and that it can prune itself from the shared tree.

Important Point to Remember

If you use a Loopback interface with a /32 mask as the RP, you must enable PIM Sparse-mode on the interface.

Configuring PIM-SM

Configuring PIM-SM is a three-step process.

1. Enable multicast routing (refer to the following step).
2. Select a rendezvous point.
3. Enable PIM-SM on an interface.

Enable multicast routing.
CONFIGURATION mode

```
ip multicast-routing
```

Related Configuration Tasks

The following are related PIM-SM configuration tasks.

- Configuring S,G Expiry Timers
- Configuring a Static Rendezvous Point
- Configuring a Designated Router
- Creating Multicast Boundaries and Domains

Enable PIM-SM

You must enable PIM-SM on each participating interface.

1. Enable multicast routing on the system.
   CONFIGURATION mode

   ```
   ip multicast-routing
   ```
2. Enable PIM-Sparse mode.
   INTERFACE mode

   ```
   ip pim sparse-mode
   ```

Examples of Viewing PIM-SM Information

To display which interfaces are enabled with PIM-SM, use the `show ip pim interface` command from EXEC Privilege mode.

```
Dell#show ip pim interface
Address     Interface VIFindex Ver/  Nbr   Query DR   DR
                               Mode  Count Intvl Prio
189.87.5.6   Te 0/11  0x2      v2/S  1     30    1    127.87.5.6
189.87.3.2   Te 0/12  0x3      v2/S  1     30    1    127.87.3.5
189.87.31.6 Te 1/11  0x0      v2/S  0     30    1    127.87.31.6
189.87.50.6 Te 1/13  0x4      v2/S  1     30    1    127.87.50.6
Dell#
```

NOTE: You can influence the selection of the Rendezvous Point by enabling PIM-Sparse mode on a Loopback interface and assigning a low IP address.

To display PIM neighbors for each interface, use the `show ip pim neighbor` command EXEC Privilege mode.

```
Dell#show ip pim neighbor
Neighbor     Interface Uptime/Expires     Ver  DR
Address                                        Prio/Mode
127.87.5.5   Te 0/11   01:44:59/00:01:16  v2   1 / S
127.87.3.5   Te 0/12   01:45:00/00:01:16  v2   1 / DR
127.87.50.5  Te 1/13   00:03:08/00:01:37  v2   1 / S
Dell#
```

To display the PIM routing table, use the `show ip pim tib` command from EXEC privilege mode.

```
Dell#show ip pim tib

PIM Multicast Routing Table
Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned,
       R - RP-bit set, F - Register flag, T - SPT-bit set, J - Join SPT,
Timers: Uptime/Expires
Interface state: Interface, next-Hop, State/Mode

(*, 192.1.2.1), uptime 00:29:36, expires 00:03:26, RP 10.87.2.6, flags: SCJ
  Incoming interface: TenGigabitEthernet 0/12, RPF neighbor 10.87.3.5
  Outgoing interface list:
    TenGigabitEthernet 0/11
    TenGigabitEthernet 1/13

(10.87.31.5, 192.1.2.1), uptime 00:01:24, expires 00:02:26, flags: FT
  Incoming interface: TenGigabitEthernet 1/11, RPF neighbor 0.0.0.0
  Outgoing interface list:
    TenGigabitEthernet 0/11
    TenGigabitEthernet 0/12
    TenGigabitEthernet 1/13
--More--
```

Configuring S,G Expiry Timers

By default, S, G entries expire in 210 seconds. You can configure a global expiry time (for all [S,G] entries) or configure an expiry time for a particular entry.
If you configure both, the ACL supersedes the global configuration for the specified entries.

When you create, delete, or update an expiry time, the changes are applied when the keep alive timer refreshes.

To configure a global expiry time or to configure the expiry time for a particular (S,G) entry, use the following commands.

1. Enable global expiry timer for S, G entries.
   CONFIGURATION mode

   ```
   ip pim sparse-mode sg-expiry-timer seconds
   ```

   The range is from 211 to 86,400 seconds.

   The default is **210**.
2. Create an extended ACL.
   CONFIGURATION mode

```
ip access-list extended access-list-name
```

3. Specify the source and group to which the timer is applied using extended ACLs with permit rules only.

   CONFIG-EXT-NACL mode

   ```
   [seq sequence-number] permit ip source-address/mask | any | host source-
   address} {destination-address/mask | any | host destination-address}
   ```

4. Set the expiry time for a specific (S,G) entry (as shown in the following example).

   CONFIGURATION mode

   ```
   ip pim sparse-mode sg-expiry-timer seconds sg-list access-list-name
   ```

   The range is from 211 to 86,400 seconds.

   The default is **210**.

Example Configuring an (S,G) Expiry Time

NOTE: The expiry time configuration is nullified and the default global expiry time is used if:

- an ACL is specified in the `ip pim sparse-mode sg-expiry-timer` command, but the ACL has not been created or is a standard ACL.
- if the expiry time is specified for an (S,G) entry in a deny rule.

```
Dell(conf)#ip access-list extended SGtimer
Dell(config-ext-nacl)#permit ip 10.1.2.3/24 225.1.1.0/24
Dell(config-ext-nacl)#permit ip any 232.1.1.0/24
Dell(config-ext-nacl)#permit ip 100.1.1.0/16 any
Dell(config-ext-nacl)#show conf
!
ip access-list extended SGtimer
  seq 5 permit ip 10.1.2.0/24 225.1.1.0/24
  seq 10 permit ip any 232.1.1.0/24
  seq 15 permit ip 100.1.0.0/16 any
Dell(config-ext-nacl)#exit
Dell(conf)#ip pim sparse-mode sg-expiry-timer 1800 sg-list SGtimer
```

To display the expiry time configuration, use the `show running-configuration [acl | pim]` command from EXEC Privilege mode.

Configuring a Static Rendezvous Point

The rendezvous point (RP) is a PIM-enabled interface on a router that acts as the root a group-specific tree; every group must have an RP.

- Identify an RP by the IP address of a PIM-enabled or Loopback interface.

  ```
  ip pim rp-address
  ```

Example of Viewing an RP on a Loopback Interface

```
Dell#sh run int loop0
!
interface Loopback 0
  ip address 1.1.1.1/32
  ip pim sparse-mode
  no shutdown
```

```
Dell#sh run pim
!
ip pim rp-address 1.1.1.1 group-address 224.0.0.0/4
```

Overriding Bootstrap Router Updates

PIM-SM routers must know the address of the RP for each group for which they have (*,G) entry.
This address is obtained automatically through the bootstrap router (BSR) mechanism or a static RP configuration.

Use the following command if you have configured a static RP for a group. If you do not use the `override` option with the following command, the RPs advertised in the BSR updates take precedence over any statically configured RPs.

- Use the `override` option to override bootstrap router updates with your static RP configuration.

  `ip pim rp-address`

Examples of Viewing the Rendezvous Point (Multicast Group) Information

To display the assigned RP for a group, use the `show ip pim rp` command from EXEC privilege mode.

```
Dell#show ip pim rp
Group        RP
225.0.1.40   165.87.50.5
226.1.1.1    165.87.50.5
```

To display the assigned RP for a group range (group-to-RP mapping), use the `show ip pim rp mapping` command in EXEC privilege mode.

```
Dell#show ip pim rp mapping
PIM Group-to-RP Mappings
Group(s): 224.0.0.0/4, Static
  RP: 165.87.50.5, v2
```

Configuring a Designated Router

Multiple PIM-SM routers might be connected to a single local area network (LAN) segment. One of these routers is elected to act on behalf of directly connected hosts. This router is the designated router (DR). The DR is elected using hello messages. Each PIM router learns about its neighbors by periodically sending a hello message out of each PIM-enabled interface. Hello messages contain the IP address of the interface out of which it is sent and a DR priority value. The router with the greatest priority value is the DR. If the priority value is the same for two routers, then the router with the greatest IP address is the DR. By default, the DR priority value is 192, so the IP address determines the DR.

- Assign a DR priority value.

  INTERFACE mode

  `ip pim dr-priority` *`priority-value`*
- Change the interval at which a router sends hello messages.

  INTERFACE mode

  `ip pim query-interval` *`seconds`*
- Display the current value of these parameter.

  EXEC Privilege mode

  `show ip pim interface`

Creating Multicast Boundaries and Domains

A PIM domain is a contiguous set of routers that all implement PIM and are configured to operate within a common boundary defined by PIM multicast border routers (PMBRs).
PMBRs connect each PIM domain to the rest of the Internet.

Create multicast boundaries and domains by filtering inbound and outbound bootstrap router (BSR) messages per interface. The following command is applied to the subsequent inbound and outbound updates. Timeout removes existing BSR advertisements.

- Create multicast boundaries and domains by filtering inbound and outbound BSR messages per interface.

  ```
  ip pim bsr-border
  ```

- Remove candidate RP advertisements.

  ```
  clear ip pim rp-mapping
  ```

Enabling PIM-SM Graceful Restart

To enable PIM-SM graceful restart, use the following commands.

- Enable PIM-SM graceful restart (non-stop forwarding capability).

  CONFIGURATION mode

  ```
  ip pim graceful-restart nsf
  ```

  - (option) `restart-time`: the time the Dell Networking system requires to restart. The default value is **180 seconds**.
  - (option) `stale-entry-time`: the maximum amount of time that the Dell Networking system preserves entries from a restarting neighbor. The default value is **60 seconds**.
  - (option) `helper-only`: this mode takes precedence over any graceful restart configuration.

  NOTE: In helper-only mode, the system preserves the PIM states of a neighboring router while the neighbor gracefully restarts, but the Dell Networking system allows itself to be taken off the forwarding path if it restarts.

Enter an example that illustrates the current task (optional).
Enter the tasks the user should do after finishing this task (optional).

34

PIM Source-Specific Mode (PIM-SSM)

PIM source-specific mode (PIM-SSM) is supported on the S6000 platform.

PIM-SSM is a multicast protocol that forwards multicast traffic from a single source to a subnet. In the other versions of protocol independent multicast (PIM), a receiver subscribes to a group only. The receiver receives traffic not just from the source in which it is interested but from all sources sending to that group. PIM-SSM requires that receivers specify the sources in which they are interested using IGMPv3 include messages to avoid receiving unwanted traffic.

PIM-SSM is more efficient than PIM-SM because it immediately creates shortest path trees (SPT) to the source rather than first using shared trees. PIM-SM requires a shared tree rooted at the RP because IGMPv2 receivers do not know about the source sending multicast data. Multicast traffic passes from the source to the receiver through the RP, until the receiver learns the source address, at which point it switches to the SPT. PIM-SSM uses IGMPv3. Because receivers subscribe to a source and group, the RP and shared tree is unnecessary; only SPTs are used. On Dell Networking systems, it is possible to use PIM-SM with IGMPv3 to achieve the same result, but PIM-SSM eliminates the unnecessary protocol overhead.

PIM-SSM also solves the multicast address allocation problem. Applications must use unique multicast addresses because if multiple applications use the same address, receivers receive unwanted traffic. However, global multicast address space is limited. Currently GLOP/EGLOP is used to statically assign Internet-routable multicast addresses, but each autonomous system number yields only 255 multicast addresses. For short-term applications, an address could be leased, but no global dynamic multicast address allocation scheme has been accepted yet. PIM-SSM eliminates the need for unique multicast addresses because routing decisions for (S1, G1) are independent from (S2, G1). As a result, subnets do not receive unwanted traffic when multiple applications use the same address.

Implementation Information

- The Dell Networking implementation of PIM-SSM is based on RFC 3569.
- The Dell Networking OS reduces the number of control messages sent between multicast routers by bundling Join and Prune requests in the same message.

Important Points to Remember

- The default SSM range is 232/8 always. Applying an SSM range does not overwrite the default range. Both the default range and SSM range are effective even when the default range is not added to the SSM ACL.
- Extended ACLs cannot be used for configuring SSM range. Be sure to create the ACL first and then apply it to the SSM range.
- The default range is always supported, so range can never be smaller than the default.

Configure PIM-SMM

Configuring PIM-SSM is a two-step process.

1. Configure PIM-SMM.
2. Enable PIM-SSM for a range of addresses.

Related Configuration Tasks

- Use PIM-SSM with IGMP Version 2 Hosts

Enabling PIM-SSM

To enable PIM-SSM, follow these steps.

1. Create an ACL that uses permit rules to specify what range of addresses should use SSM.
   CONFIGURATION mode

   ```
   ip access-list standard name
   ```
2. Enter the `ip pim ssm-range` command and specify the ACL you created.
   CONFIGURATION mode

   ```
   ip pim ssm-range acl-name
   ```

Enabling PIM-SSM

To display address ranges in the PIM-SSM range, use the `show ip pim ssm-range` command from EXEC Privilege mode.

```
R1(conf)#do show run pim
!
ip pim rp-address 10.11.12.2 group-address 224.0.0.0/4
ip pim ssm-range ssm
R1(conf)#do show run acl
!
ip access-list standard ssm
  seq 5 permit host 239.0.0.2
R1(conf)#do show ip pim ssm-range
Group Address   / MaskLen
239.0.0.2       / 32
```

Use PIM-SSM with IGMP Version 2 Hosts

PIM-SSM requires receivers that support IGMP version 3. You can employ PIM-SSM even when receivers support only IGMP version 1 or version 2 by translating (*,G) entries to (S,G) entries.

Translate (*,G) entries to (S,G) entries using the `ip igmp ssm-map acl` command source from CONFIGURATION mode. In a standard access list, specify the groups or the group ranges that you want to map to a source. Then, specify the multicast source.

- When an SSM map is in place and Dell Networking OS cannot find any matching access lists for a group, it continues to create (*,G) entries because there is an implicit deny for unspecified groups in the ACL.

- When you remove the mapping configuration, Dell Networking OS removes the corresponding (S,G) states that it created and re-establishes the original (*,G) states.
- You may enter multiple `ssm-map` commands for different access lists. You may also enter multiple `ssm-map` commands for the same access list, as long as they use different source addresses.
- When an extended ACL is associated with this command, Dell Networking OS displays an error message. If you apply an extended ACL before you create it, Dell Networking OS accepts the configuration, but when the ACL is later defined, Dell Networking OS ignores the ACL and the stated mapping has no effect.

To display the source to which a group is mapped, use the `show ip igmp ssm-map [group]` command. If you use the `group` option, the command displays the group-to-source mapping even if the group is not currently in the IGMP group table. If you do not specify the `group` option, the display is a list of groups currently in the IGMP group table that has a group-to-source mapping.

To display the list of sources mapped to a group currently in the IGMP group table, use the `show ip igmp groups group detail` command.

Configuring PIM-SSM with IGMPv2

```
R1(conf)#do show run pim
!
ip pim rp-address 10.11.12.2 group-address 224.0.0.0/4
ip pim ssm-range ssm
R1(conf)#do show run acl
!
ip access-list standard map
seq 5 permit host 239.0.0.2
!
ip access-list standard ssm
  seq 5 permit host 239.0.0.2
R1(conf)#ip igmp ssm-map map 10.11.5.2
R1(conf)#do show ip igmp groups
Total Number of Groups: 2
IGMP Connected Group Membership
Group Address   Interface   Mode          Uptime    Expires  Last Reporter
239.0.0.2       Vlan 300    IGMPv2-Compat 00:00:07  Never    10.11.3.2
   Member Ports: Gi 1/1
239.0.0.1 Vlan 400 INCLUDE 00:00:10 Never 10.11.4.2
R1(conf)#do show ip igmp ssm-map
IGMP Connected Group Membership
Group Address  Interface   Mode           Uptime    Expires  Last Reporter
239.0.0.2      Vlan 300    IGMPv2-Compat  00:00:36  Never    10.11.3.2
   Member Ports: Gi 1/1
R1(conf)#do show ip igmp ssm-map 239.0.0.2
SSM Map Information
Group     : 239.0.0.2
Source(s) : 10.11.5.2
R1(conf)#do show ip igmp groups detail

Interface           Vlan 300
Group               239.0.0.2
Uptime              00:00:01
Expires             Never
Router mode         IGMPv2-Compat
Last reporter       10.11.3.2
Last reporter mode  IGMPv2
Last report         received Join
Group source        list
Source address      Uptime Expires
10.11.5.2 00:00:01  Never
```

```
Interface             Vlan 400
Group                 239.0.0.1
Uptime                00:00:05
Expires               Never
Router mode           INCLUDE
Last reporter         10.11.4.2
Last reporter mode    INCLUDE
Last report received  ALLOW
Group source list
Source address  Uptime    Expires
10.11.5.2       00:00:05  00:02:04
  Member Ports: Gi 1/2
```

35

Port Monitoring

Port monitoring is supported on the S6000 platform.

Mirroring is used for monitoring Ingress or Egress or both Ingress and Egress traffic on a specific port(s). This mirrored traffic can be sent to a port where a network sniffer can connect and monitor the traffic.

Dell Networking OS supports the following mirroring techniques:

- Port-Mirroring — Port Monitoring is a method of monitoring network traffic that forwards a copy of each incoming and outgoing packet from one port of a network router to another port where the packet can be studied.
- Remote Port Monitoring (RPM) — Remote Port Monitoring allows the user to monitor traffic running across a remote device in the same network. Here the mirror traffic is carried over the L2 network, so that probe devices in the network can analyze it. It is an extension to the normal Port Monitoring feature. This feature is generally referred as RPM, where mirror traffic is carried over L2 network.
- Encapsulated Remote-Port Monitoring (ERPM) — ERPM is a feature to encapsulate mirrored packet using GRE with IP delivery so that it can be sent across a routed network.

Important Points to Remember

- Port Monitoring is supported on both physical and logical interfaces like virtual area network (VLAN) and port-channel.
- The monitored (the source, [MD]) and monitoring ports (the destination, [MG]) must be on the same switch.
- In general, a monitoring port should have no ip address and no shutdown as the only configuration; Dell Networking OS permits a limited set of commands for monitoring ports. You can display these commands using the `?` command. A monitoring port also may not be a member of a VLAN.
- There may only be one destination port (MG) in a monitoring session.
- Source port (MD) can be monitored by more than one destination port (MG).
- Destination port (MG) can be a physical interface or port-channel interface.
- A Port monitoring session can have multiple source statements.
- Range command is supported in the source statement, where we can specify a range of interfaces of (Physical, Port Channel or VLAN) types.
- One Destination Port (MG) can be used in multiple sessions.
- There can be a maximum of 128 source ports in a Port Monitoring session.
- Flow based monitoring is supported for all type of source interfaces.
- Source port (MD) can be a VLAN, where the VLAN traffic received on that port pipe where its members are present is monitored
- Single MD can be monitored on max. of 4 MG ports.

Port Monitoring

The S6000 supports multiple source-destination statements in a single monitor session.

The maximum number of source ports that can be supported in a session is 128.

The maximum number of destination ports that can be supported is 4 per port pipe.

In the following examples, ports 0/13, 0/14, 0/15, and 0/16 all belong to the same port-pipe. They are pointing to four different destinations (0/1, 0/2, 0/3, and 0/37). Now it is not possible for another source port from the same port-pipe (for example, 0/17) to point to another new destination (for example, 0/4). If you attempt to configure another destination (to create 5 MG port), this message displays: `% Error will be thrown in case of RPM and ERPM features`.

Example of Changing the Destination Port in a Monitoring Session

```
Dell(conf-mon-sess-5)#do show moni session
  SessID  Source         Destination         Dir  Mode  Source IP      Dest IP
  ------  ------         -----------         ---  ----  ---------      --------
     1   Te 0/0            Te 0/1            both Port     N/A            N/A
     2   Te 0/0            Te 0/2            both Port     N/A            N/A
     3   Te 0/0            Te 0/3            both Port     N/A            N/A
     4   Te 0/0            Te 0/4            both Port     N/A            N/A
     5   Te 0/0            Te 0/5            both Port     N/A            N/A
Dell(conf-mon-sess-5)#

Dell(conf)#mon ses 300
Dell(conf-mon-sess-300)#source tegig 0/17 destination tegig 0/4 direction tx
% Error: Exceeding max MG ports for this MD port pipe.
Dell(conf-mon-sess-300)#
Dell(conf-mon-sess-300)#source tegig 0/17 destination tegig 0/1 direction tx
Dell(conf-mon-sess-300)#do show mon session
SessionID Source  Destination Direction Mode      Type
--------- ------  ----------- --------- ----      ----
0         Te 0/13 Gi 0/1      rx        interface Port-based
10        Te 0/14 Gi 0/2      rx        interface Port-based
20        Te 0/15 Gi 0/3      rx        interface Port-based
30        Te 0/16 Gi 0/37     rx        interface Port-based
300       Te 0/17 Gi 0/1      tx        interface Port-based
Dell(conf-mon-sess-300)#
```

Example of Configuring Another Monitoring Session with a Previously Used Destination Port

```
Dell(conf)#mon ses 300
Dell(conf-mon-sess-300)#source TeGig 0/17 destination Te    Gig 0/4 direction tx
% Error: Exceeding max MG ports for this MD port pipe.
Dell(conf-mon-sess-300)#
Dell(conf-mon-sess-300)#source TeGig 0/17 destination TeGig 0/1 direction tx
Dell(conf-mon-sess-300)#do show mon session
SessionID Source  Destination Direction Mode      Type
--------- ------  ----------- --------- ----      ----
0         Te 0/13 Gi 0/1      rx        interface Port-based
10        Te 0/14 Gi 0/2      rx        interface Port-based
20        Te 0/15 Gi 0/3      rx        interface Port-based
30        Te 0/16 Gi 0/37     rx        interface Port-based
300       Te 0/17 Gi 0/1      tx        interface Port-based
```

Example of Viewing a Monitoring Session

In the example below, 0/25 and 0/26 belong to Port-pipe 1. This port-pipe has the same restriction of only four destination ports, new or used.

```
Dell(conf-mon-sess-300)#do show mon session
SessionID Source  Destination Direction Mode      Type
--------- ------  ----------- --------- ----      ----
0         Te 0/13 Gi 0/1      rx        interface Port-based
10        Te 0/14 Gi 0/2      rx        interface Port-based
20        Te 0/15 Gi 0/3      rx        interface Port-based
30        Te 0/16 Gi 0/37     rx        interface Port-based
100       Te 0/25 Gi 0/38     tx        interface Port-based
110       Te 0/26 Gi 0/39     tx        interface Port-based
300       Te 0/17 Gi 0/1      tx        interface Port-based
Dell(conf-mon-sess-300)#
```

Given these parameters, the following illustration shows the possible port monitoring configurations on the S-Series.

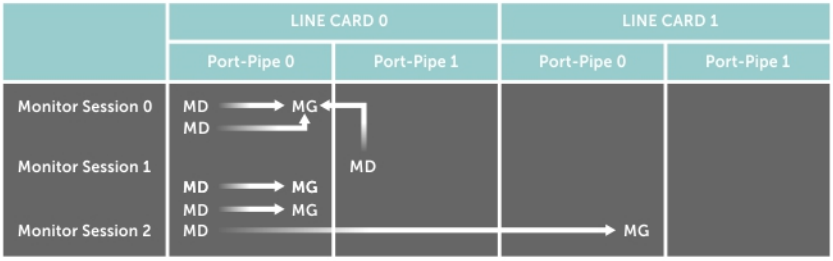


Figure 94. Port Monitoring Configurations on the S-Series

Dell Networking OS Behavior: All monitored frames are tagged if the configured monitoring direction is egress (TX), regardless of whether the monitored port (MD) is a Layer 2 or Layer 3 port. If the MD port is a Layer 2 port, the frames are tagged with the VLAN ID of the VLAN to which the MD belongs. If the MD port is a Layer 3 port, the frames are tagged with VLAN ID 4095. If the MD port is in a Layer 3 VLAN, the frames are tagged with the respective Layer 3 VLAN ID. For example, in the configuration source TeGig 6/0 destination TeGig 6/1 direction tx, if the MD port TenGig 6/0 is an untagged member of any VLAN, all monitored frames that the MG port TeGig 6/1 receives are tagged with the VLAN ID of the MD port. Similarly, if BPDUs are transmitted, the MG port receives them tagged with the VLAN ID 4095. This behavior might result in a difference between the number of egress packets on the MD port and monitored packets on the MG port.

Dell Networking OS Behavior: The continues to mirror outgoing traffic even after an MD participating in spanning tree protocol (STP) transitions from the forwarding to blocking.

Configuring Port Monitoring

To configure port monitoring, use the following commands.

1. Verify that the intended monitoring port has no configuration other than no shutdown, as shown in the following example.

   EXEC Privilege mode

```
show interface
```

2. Create a monitoring session using the command monitor session from CONFIGURATION mode, as shown in the following example.

   CONFIGURATION mode

   ```
   monitor session
   ```

   ```
   monitor session type rpm/erpm
   type is an optional keyword, required only for rpm and erpm
   ```

3. Specify the source and destination port and direction of traffic, as shown in the following example.

   MONITOR SESSION mode

   ```
   source
   ```

Example of Viewing Port Monitoring Configuration

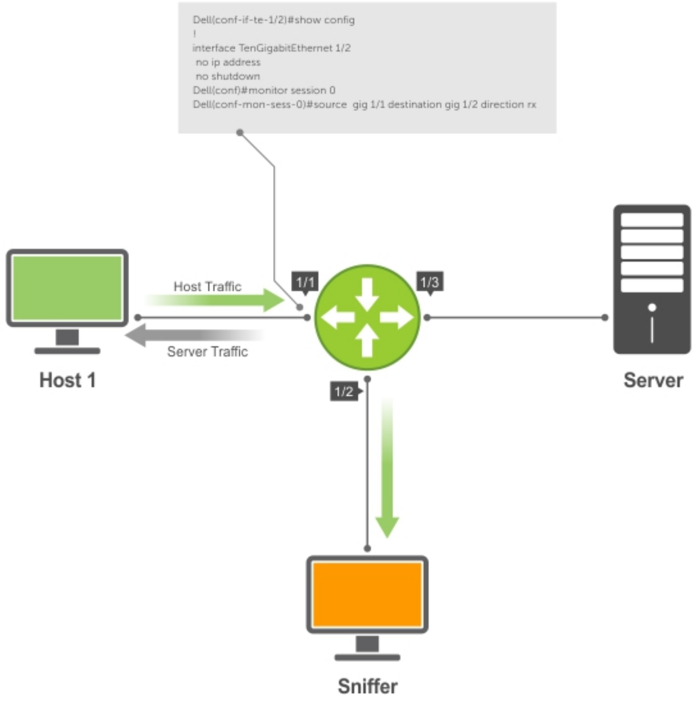
To display information on currently configured port-monitoring sessions, use the `show monitor session` command from EXEC Privilege mode.

```
Dell(conf)#monitor session 0
Dell(conf-mon-sess-0)#$source ten 0/0 dest ten 0/1 dir rx
Dell(conf-mon-sess-0)#show c
!
monitor session 0
 source TenGigabitEthernet 0/0 destination TenGigabitEthernet 0/1 direction rx
Dell(conf-mon-sess-0)#
Dell(conf-mon-sess-0)#do show monitor session
  SessID  Source    Destination    Dir  Mode  Source IP  Dest IP
  ------  ------    ------------   ---  ----  ---------   --------
     0   Te 0/0       Te 0/1       rx   Port     N/A         N/A

Dell(conf)#monitor session 0
Dell(conf-mon-sess-0)#source po 10 dest ten 0/1 dir rx
Dell(conf-mon-sess-0)#do show monitor session
  SessID  Source   Destination     Dir  Mode  Source IP    Dest IP
  ------  ------   -----------     ---  ----  ---------    --------
     0   Te 0/0       Te 0/1        rx   Port     N/A        N/A
     0   Po 10        Te 0/1        rx   Port     N/A        N/A

Dell(conf)#monitor session 1
Dell(conf-mon-sess-1)#source vl 40 dest ten 0/2 dir rx
Dell(conf-mon-sess-1)#flow-based enable
Dell(conf-mon-sess-1)#exit
Dell(conf)#do show monitor session
  SessID  Source   Destination   Dir  Mode  Source IP      Dest IP
  ------  ------   -----------   ---  ----  ---------      --------
     0   Te 0/0      Te 0/1       rx   Port     N/A            N/A
     0   Po 10       Te 0/1       rx   Port     N/A            N/A
     1   Vl 40       Te 0/2       rx   Flow     N/A            N/A
```

Note: Source as VLAN is achieved via Flow based mirroring. Please refer section Enabling Flow-Based Monitoring.

In the following example, the host and server are exchanging traffic which passes through the uplink interface 1/1. Port 1/1 is the monitored port and port 1/42 is the destination port, which is configured to only monitor traffic received on tengigabitethernet 1/1 (host-originated traffic).

Figure 95. Port Monitoring Example



```
Dell(conf-if-te-1/2)#show config
!
interface TenGigabitEthernet 1/2
 no ip address
 no shutdown
Dell(conf)#monitor session 0
Dell(conf-mon-sess-0)#source  gig 1/1 destination gig 1/2 direction rx
```

Enabling Flow-Based Monitoring

Flow-based monitoring is supported only on the S-Series platform.
Flow-based monitoring conserves bandwidth by monitoring only specified traffic instead of all traffic on the interface. This feature is particularly useful when looking for malicious traffic. It is available for Layer 2 and Layer 3 ingress and egress traffic. You can specify traffic using standard or extended access-lists.

1. Enable flow-based monitoring for a monitoring session.

   MONITOR SESSION mode

   ```
   flow-based enable
   ```
2. Define in access-list rules that include the keyword `monitor`. For port monitoring, Dell Networking OS only considers traffic matching rules with the keyword `monitor`.

   CONFIGURATION mode

   ```
   ip access-list
   ```

   Refer to Access Control Lists (ACLs).
3. Apply the ACL to the monitored port.

   INTERFACE mode

   ```
   ip access-group access-list
   ```

Example of the `flow-based enable` Command

To view an access-list that you applied to an interface, use the `show ip accounting access-list` command from EXEC Privilege mode.

```
Dell(conf)#monitor session 0
Dell(conf-mon-sess-0)#flow-based enable
Dell(conf)#ip access-list ext testflow
Dell(config-ext-nacl)#seq 5 permit icmp any any count bytes monitor
Dell(config-ext-nacl)#seq 10 permit ip 102.1.1.0/24 any count bytes monitor
Dell(config-ext-nacl)#seq 15 deny udp any any count bytes
Dell(config-ext-nacl)#seq 20 deny tcp any any count bytes
Dell(config-ext-nacl)#exit
Dell(conf)#interface gig 1/1
Dell(conf-if-gi-1/1)#ip access-group testflow in
Dell(conf-if-gi-1/1)#show config
!
interface GigabitEthernet 1/1
  ip address 10.11.1.254/24
  ip access-group testflow in
  shutdown
Dell(conf-if-gi-1/1)#exit
Dell(conf)#do show ip accounting access-list testflow
!
Extended Ingress IP access list testflow on GigabitEthernet 1/1
Total cam count 4
  seq 5 permit icmp any any monitor count bytes (0 packets 0 bytes)
  seq 10 permit ip 102.1.1.0/24 any monitor count bytes (0 packets 0 bytes)
  seq 15 deny udp any any count bytes (0 packets 0 bytes)
  seq 20 deny tcp any any count bytes (0 packets 0 bytes)
Dell(conf)#do show monitor session 0
SessionID Source Destination Direction Mode      Type
--------- ------ ----------- --------- ----      ----
0         Gi 1/1 Gi 1/2      rx        interface Flow-based
```

Remote Port Mirroring

While local port monitoring allows you to monitor traffic from one or more source ports by directing it to a destination port on the same switch/router, remote port mirroring allows you to monitor Layer 2 and Layer 3 ingress and/or egress traffic on multiple source ports on different switches and forward the mirrored traffic to multiple destination ports on different switches.

Remote port mirroring helps network administrators monitor and analyze traffic to troubleshoot network problems in a time-saving and efficient way.

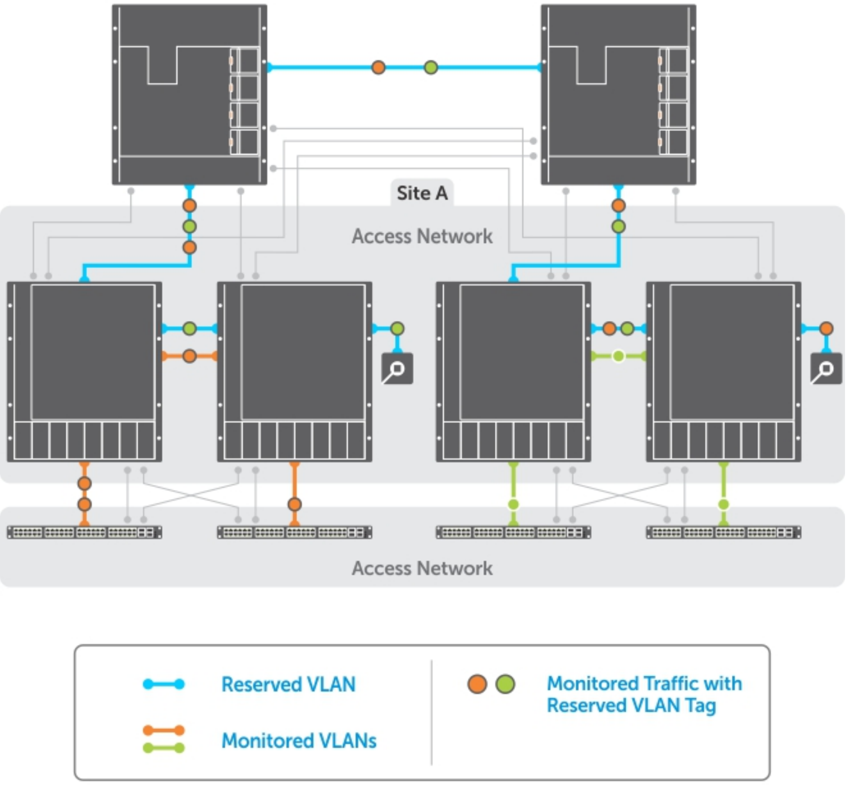
In a remote-port mirroring session, monitored traffic is tagged with a VLAN ID and switched on a user-defined, non-routable L2 VLAN. The VLAN is reserved in the network to carry only mirrored traffic, which is forwarded on all egress ports of the VLAN. Each intermediate switch that participates in the transport of mirrored traffic must be configured with the reserved L2 VLAN. Remote port monitoring supports mirroring sessions in which multiple source and destination ports are distributed across multiple switches

Remote Port Mirroring Example

Remote port mirroring uses the analyzers shown in the aggregation network in Site A.

The VLAN traffic on monitored links from the access network is tagged and assigned to a dedicated L2 VLAN. Monitored links are configured in two source sessions shown with orange and green circles. Each source session uses a separate reserved VLAN to transmit mirrored packets (mirrored source-session traffic is shown with an orange or green circle with a blue border).

The reserved VLANs transport the mirrored traffic in sessions (blue pipes) to the destination analyzers in the local network. Two destination sessions are shown: one for the reserved VLAN that transports orange-circle traffic; one for the reserved VLAN that transports green-circle traffic.



Configuring Remote Port Mirroring

Remote port mirroring requires a source session (monitored ports on different source switches), a reserved tagged VLAN for transporting mirrored traffic (configured on source, intermediate, and destination switches), and a destination session (destination ports connected to analyzers on destination switches).

Configuration Notes

When you configure remote port mirroring, the following conditions apply:

- You can configure any switch in the network with source ports and destination ports, and allow it to function in an intermediate transport session for a reserved VLAN at the same time for multiple remote-port mirroring sessions. You can enable and disable individual mirroring sessions.
- BPDU monitoring is not required to use remote port mirroring.
- A remote port mirroring session mirrors monitored traffic by prefixing the reserved VLAN tag to monitored packets so that they are copied to the reserve VLAN.

- Mirrored traffic is transported across the network using 802.1Q-in-802.1Q tunneling. The source address, destination address and original VLAN ID of the mirrored packet are preserved with the tagged VLAN header. Untagged source packets are tagged with the reserve VLAN ID.
- The RPM VLAN can't be a Private VLAN.
- The RPM VLAN can be used as GVRP VLAN.
- The L3 interface configuration should be blocked for RPM VLAN.
- The member port of the reserved VLAN should have MTU and IPMTU value as MAX+4 (to hold the VLAN tag parameter).
- To associate with source session, the reserved VLAN can have at max of only 4 member ports.
- To associate with destination session, the reserved VLAN can have multiple member ports.
- Reserved Vlan cannot have untagged ports

In the reserved **L2 VLAN** used for remote port mirroring:

- MAC address learning in the reserved VLAN is automatically disabled.
- The reserved VLAN for remote port mirroring can be automatically configured in intermediate switches by using GVRP.
- There is no restriction on the VLAN IDs used for the reserved remote-mirroring VLAN. Valid VLAN IDs are from 2 to 4094. The default VLAN ID is not supported.
- In mirrored traffic, packets that have the same destination MAC address as an intermediate or destination switch in the path used by the reserved VLAN to transport the mirrored traffic are dropped by the switch that receives the traffic if the switch has a L3 VLAN configured.

In a **source session** used for remote port mirroring:

- Maximum number of source sessions supported on a switch: 4
- Maximum number of source ports supported in a source session: 128
- You can configure physical ports and port-channels as sources in remote port mirroring and use them in the same source session. You can use both Layer 2 (configured with the switchport command) and Layer 3 ports as source ports. You can optionally configure one or more source VLANs to specify the VLAN traffic to be mirrored on source ports.
- You can use the default VLAN and native VLANs as a source VLAN.
- You cannot configure the dedicated VLAN used to transport mirrored traffic as a source VLAN.
- Egressing remote-vlan packets are rate limited to a default value of 100 Mbps.

In a **destination session** used for remote port mirroring:

- Maximum number of destination sessions supported on a switch: 64
- Maximum number ports supported in a destination session: 64.
- You can configure any port as a destination port.
- You can configure additional destination ports in an active session.
- You can tunnel the mirrored traffic from multiple remote-port source sessions to the same destination port.
- By default, destination port sends the mirror traffic to the probe port by stripping off the rpm header. We can also configure the destination port to send the mirror traffic with the rpm header intact in the original mirror traffic..
- By default, ingress traffic on a destination port is dropped.

Restrictions

When you configure remote port mirroring, the following **restrictions** apply:

- You can configure the same source port to be used in multiple source sessions.
- You cannot configure a source port channel or source VLAN in a source session if the port channel or VLAN has a member port that is configured as a destination port in a remote-port mirroring session.
- A destination port for remote port mirroring cannot be used as a source port, including the session in which the port functions as the destination port.
- A destination port cannot be used in any spanning tree instance.
- The reserved VLAN used to transport mirrored traffic must be a L2 VLAN. L3 VLANs are not supported.
- On a source switch on which you configure source ports for remote port mirroring, you can add only one port to the dedicated RPM VLAN which is used to transport mirrored traffic. You can configure multiple ports for the dedicated RPM VLAN on intermediate and destination switches.

Displaying Remote-Port Mirroring Configurations

To display the current configuration of remote port mirroring for a specified session, enter the **show config** command in **MONITOR SESSION** configuration mode.

```
Dell(conf-mon-sess-2)#show config
!
monitor session 2 type rpm
 source fortyGigE 0/60 destination remote-vlan 300 direction rx
 source Port-channel 10 destination remote-vlan 300 direction rx
 no disable
```

To display the currently configured source and destination sessions for remote port mirroring on a switch, enter the **show monitor session** command in **EXEC** Privilege mode.

```
Dell(conf)#do show monitor session
  SessID  Source         Destination         Dir  Mode  Source IP      Dest IP
  ------  ------         -----------         ---  ----  ---------      --------
     1   remote-vlan 100   Fo 0/48           N/A  N/A      N/A            N/A
     1   remote-vlan 100   Po 100            N/A  N/A      N/A            N/A
     2   Fo 0/60           remote-vlan 300   rx   Port     N/A            N/A
     2   Po 10             remote-vlan 300   rx   Port     N/A            N/A
```

To display the current configuration of the reserved VLAN, enter the **show vlan** command.

```
Dell#show vlan

Codes: * - Default VLAN, G - GVRP VLANs, R - Remote Port Mirroring VLANs, P -
Primary, C - Community, I - Isolated
       O - Openflow
Q: U - Untagged, T - Tagged
   x - Dot1x untagged, X - Dot1x tagged
   o - OpenFlow untagged, O - OpenFlow tagged
   G - GVRP tagged, M - Vlan-stack
   i - Internal untagged, I - Internal tagged, v - VLT untagged, V - VLT tagged

    NUM    Status    Description                     Q Ports
*   1      Inactive
R   100    Active                                   T Fo 0/44
R   300    Active                                   T Fo 0/52
```

Configuring the Sample Remote Port Mirroring

Remote port mirroring requires a source session (monitored ports on different source switches), a reserved tagged VLAN for transporting mirrored traffic (configured on source, intermediate, and destination switches), and a destination session (destination ports connected to analyzers on destination switches).

Configuration Steps for RPM

Step	Command	Purpose
1	configure terminal	Enter global configuration mode.
2	**monitor session** <id> **type rpm**	The <id> needs to be unique and not already defined in the box specifying type as 'rpm' defines a RPM session.
3	**source** *Interface* | *Range*	Specify the port or list of ports that needs to be monitored
4	**direction**	Specify rx, tx or both in case to monitor ingress/ egress or both ingress and egress packets on the specified port..
5	**rpm source-ip <id>** ***dest-ip*** **<id>**	Specify the source ip address and the destination ip where the packet needs to be sent.
6	**<no>** ***flow-based enable***	Specify flow-based enable for mirroring on a flow by flow basis and also for vlan as source.
7	***no enable***	(Optional) No disable command is mandatory in order for a rpm session to be active.

Configuring the sample Source Remote Port Mirroring

```
Dell(conf)#interface vlan 10
Dell(conf-if-vl-10)#mode remote-port-mirroring
Dell(conf-if-vl-10)#tagged te 0/4
Dell(conf-if-vl-10)#exit

Dell(conf)#monitor session 1 type rpm
Dell(conf-mon-sess-1)#source te 0/5 destination remote-vlan 10 dir rx
Dell(conf-mon-sess-1)#no disable
Dell(conf-mon-sess-1)#exit

Dell(conf)#inte vlan 100
Dell(conf-if-vl-100)#tagged te 0/7
Dell(conf-if-vl-100)#exit

Dell(conf)#interface vlan 20
Dell(conf-if-vl-20)#mode remote-port-mirroring
Dell(conf-if-vl-20)#tagged te 0/6
Dell(conf-if-vl-20)#exit

Dell(conf)#monitor session 2 type rpm
Dell(conf-mon-sess-2)#source vlan 100 destination remote-vlan 20 dir rx
Dell(conf-mon-sess-2)#no disable
Dell(conf-mon-sess-2)#flow-based enable
Dell(conf-mon-sess-2)#exit

Dell(conf)#mac access-list standard mac_acl
Dell(config-std-macl)#permit 00:00:00:00:11:22 count monitor
Dell(config-std-macl)#exit

Dell(conf)#interface vlan 100
Dell(conf-if-vl-100)#mac access-group mac_acl1 in
Dell(conf-if-vl-100)#exit
Dell(conf)#inte te 0/30
Dell(conf-if-te-0/30)#no shutdown
Dell(conf-if-te-0/30)#switchport
Dell(conf-if-te-0/30)#exit

Dell(conf)#interface vlan 30
Dell(conf-if-vl-30)#mode remote-port-mirroring
Dell(conf-if-vl-30)#tagged te 0/30
Dell(conf-if-vl-30)#exit

Dell(conf)#interface port-channel 10
Dell(conf-if-po-10)#channel-member te 0/28-29
Dell(conf-if-po-10)#no shutdown
Dell(conf-if-po-10)#exit

Dell(conf)#monitor session 3 type rpm
Dell(conf-mon-sess-3)#source port-channel 10 dest remote-vlan 30 dir both
Dell(conf-mon-sess-3)#no disable
Dell(conf-mon-sess-3)#
Dell(conf-mon-sess-3)#exit
Dell(conf)#end
Dell#

Dell#show monitor session
  SessID  Source         Destination         Dir  Mode  Source IP      Dest IP
  ------  ------         -----------         ---  ----  ---------      --------
     1   Te 0/5            remote-vlan 10    rx   Port     N/A            N/A
     2   Vl 100            remote-vlan 20    rx   Flow     N/A            N/A
     3   Po 10             remote-vlan 30    both Port     N/A            N/A
Dell#
```

Configuring the sample Source Remote Port Mirroring

```
Dell(conf)#inte te 0/0
Dell(conf-if-te-0/0)#switchport
Dell(conf-if-te-0/0)#no shutdown
Dell(conf-if-te-0/0)#exit

Dell(conf)#interface te 0/1
Dell(conf-if-te-0/1)#switchport
Dell(conf-if-te-0/1)#no shutdown
Dell(conf-if-te-0/1)#exit

Dell(conf)#interface te 0/2
Dell(conf-if-te-0/2)#switchport
Dell(conf-if-te-0/2)#no shutdown
Dell(conf-if-te-0/2)#exit

Dell(conf)#inte vlan 10
Dell(conf-if-vl-10)#mode remote-port-mirroring
Dell(conf-if-vl-10)#tagged te 0/0
Dell(conf-if-vl-10)#exit

Dell(conf)#inte vlan 20
Dell(conf-if-vl-20)#mode remote-port-mirroring
Dell(conf-if-vl-20)#tagged te 0/1
Dell(conf-if-vl-20)#exit

Dell(conf)#interface vlan 30
Dell(conf-if-vl-30)#mode remote-port-mirroring
Dell(conf-if-vl-30)#tagged te 0/2
Dell(conf-if-vl-30)#exit
Dell(conf)#monitor session 1 type rpm
Dell(conf-mon-sess-1)#source remote-vlan 10 dest te 0/3
Dell(conf-mon-sess-1)#exit

Dell(conf)#monitor session 2 type rpm
Dell(conf-mon-sess-2)#source remote-vlan 20 destination te 0/4
Dell(conf-mon-sess-2)#tagged destination te 0/4
Dell(conf-mon-sess-2)#exit

Dell(conf)#monitor session 3 type rpm
Dell(conf-mon-sess-3)#source remote-vlan 30 destination te 0/5
Dell(conf-mon-sess-3)#tagged destination te 0/5
Dell(conf-mon-sess-3)#end
Dell#
Dell#show monitor session
  SessID  Source         Destination         Dir  Mode  Source IP      Dest IP
  ------  ------         -----------         ---  ----  ---------      --------
     1   remote-vlan 10    Te 0/3            N/A  N/A      N/A            N/A
     2   remote-vlan 20    Te 0/4            N/A  N/A      N/A            N/A
     3   remote-vlan 30    Te 0/5            N/A  N/A      N/A            N/A
Dell#
```

Configuring RSPAN Source Sessions to Avoid BPD Issues

When ever you configure an RPM source session, you must ensure the following to avoid BPDU issues:

1. Enable control plane egress acl using the following command:

   ```
   mac control-plane egress-acl
   ```

2. Create an extended MAC access list and add a deny rule of (0x0180c2xxxxxx) packets using the following commands:

   ```
   mac access-list extended mac2
   seq 5 deny any 01:80:c2:00:00:00 00:00:00:ff:ff:ff count
   ```

3. Apply ACL on that RPM VLAN. In this example RPM vlan is 10.

   ```
   Dell#show running-config interface vlan 10
   !
   interface Vlan 10
   no ip address
   mode remote-port-mirroring
   tagged Port-channel 2
   mac access-group mac2 out
   no shutdown
   ```

4. Create Source RPM session as follows (port-channel 1 and port-channel 2 are LACP).

   ```
       Dell(conf)#monitor session 1 type rpm
       Dell(conf-mon-sess-1)#source port-channel 1 destination remote-vlan 10
   dir rx
       Dell(conf-mon-sess-1)#no disable
   ```

5. Show the output for the LACP.

   ```
       Dell#show interfaces port-channel brief
       Codes: L - LACP Port-channel
       O - OpenFlow Controller Port-channel

       LAG  Mode  Status       Uptime      Ports
       L1   L3    up           00:01:17    Te 0/44    (Up)
       L2   L2    up           00:00:58    Te 0/45    (Up)
   Dell#
   ```

Configuring the Encapsulated Remote Port Mirroring

The ERPM session copies traffic from the source ports/lags or source VLANs and forwards the traffic using routable GRE-encapsulated packets to the destination ip address specified in the session.

Important:

The steps to be followed for the ERPM Encapsulation :

- Dell Networking OS supports ERPM Source session only. The Encapsulated packets terminate at the destination ip or at the analyzer.
- Make sure that the destination ip is reachable via the configured ip route (static or dynamic)
- The system MTU should be configured properly to accommodate the increased size of the ERPM mirrored packet.
- The system encapsulates the complete ingress or egress data under GRE header, IP header and outer MAC header and sends it out at the next hop interface as pointed by the routing table.
- The source IP address can be any port's ip address defined in the box but it should be unique and should not be assigned to any other system in the network.
- The keyword ‘flow-based enable’ should have been specified in case of source as vlan or where monitoring on a per flow basis is desired.
- The keyword monitor should have been specified in the access list rules for which we need to mirror.
- The maximum number of source ports that can be defined in a session is 128.
- The system allows to configure upto 4 ERPM sessions.
- ERPM sessions do not copy locally sourced Remote-VLAN traffic from source trunk ports that carry RPM VLANs. ERPM sessions do not copy locally sourced ERPM GRE-encapsulated traffic from source ports.
- Source Vlan monitoring can be done only for ingress packets and is not supported for egress direction.

Configuration steps for ERPM

ERPM sample steps for configuring the Source sessions

Step	Command	Purpose
1	configure terminal	Enter global configuration mode.
2	**monitor session** <id> **type erpm**	The <id> needs to be unique and not already defined in the box specifying type as 'erpm' defines a RPM session.
3	**source** *Interface* | *Range*	Specify the port or list of ports that needs to be monitored
4	**direction**	Specify rx, tx or both in case to monitor ingress/egress or both ingress and egress packets on the specified port..
5	**erpm source-ip <id> *dest-ip* <id>**	Specify the source ip address and the destination ip where the packet needs to be sent.

6	<no> *flow-based enable*	Specify flow-based enable for mirroring on a flow by flow basis and also for vlan as source.
7	*no enable*	(Optional) No disable command is mandatory in order for a erpm session to be active.

The following example shows a sample configuration .

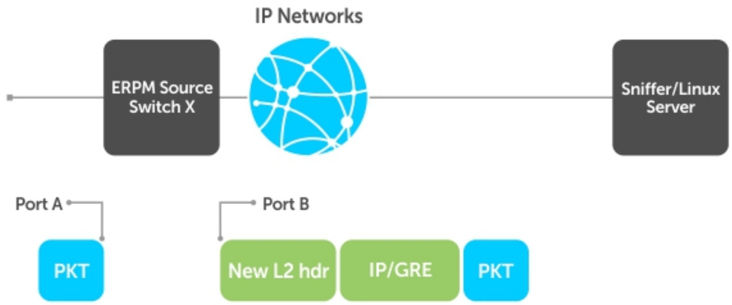
```
Dell(conf)#monitor session 0 type erpm
Dell(conf-mon-sess-0)#source tengigabitethernet 0/9 direction rx
Dell(conf-mon-sess-0)#source port-channel 1 direction tx
Dell(conf-mon-sess-0)#erpm source-ip 1.1.1.1 dest-ip 7.1.1.2
Dell(conf-mon-sess-0)#no disable

Dell(conf)#monitor session 1 type erpm
Dell(conf-mon-sess-1)#source vlan 11 direction rx
Dell(conf-mon-sess-1)#erpm source-ip 5.1.1.1 dest-ip 3.1.1.2
Dell(conf-mon-sess-1)#flow-based enable
Dell(conf-mon-sess-1)#no disable

Dell# show monitor session
SessID Source   Destination Dir Mode Source IP Dest IP
 ------ ------  ----------- --- ---- --------- --------
0       Te 0/9  remote-ip  rx  Port   1.1.1.1 7.1.1.2
0       Po 1    remote-ip  tx  Port   1.1.1.1 7.1.1.2
1       Vl 11   remote-ip  rx  Flow   5.1.1.1 3.1.1.2
```

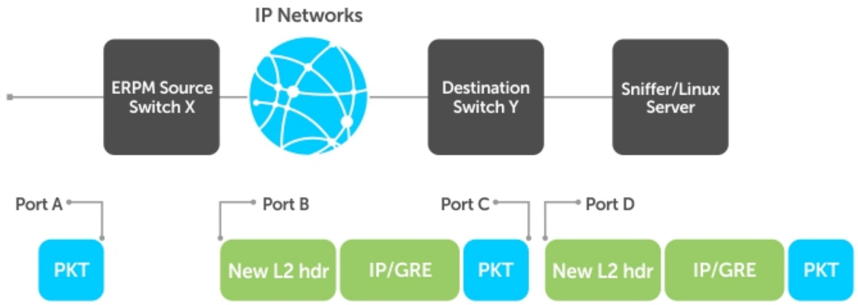
Sample example for monitoring the VLANs as source, an access list with monitor keyword in its rules needs to be attached to the vlan interface.

```
Dell(conf)#mac access-list standard flow
Dell(config-std-macl)#seq 5 permit 00:00:0a:00:00:0b count monitor

Dell#show running-config interface vlan 11
!
interface Vlan 11
 no ip address
 tagged TenGigabitEthernet 0/1-3
 mac access-group flow in  <<<<<<<<<<<<<< Only ingress packets are supported
for mirroring
 shutdown
Dell#
```

ERPM Behavior on a typical Dell Networking OS

The Dell Networking OS is designed to support only the Encapsulation of the data received / transmitted at the specified source port (Port A). An ERPM destination session / decapsulation of the ERPM packets at the destination Switch are not supported.


IP Networks
ERPM Source Switch X
Sniffer/Linux Server
Port A
Port B
PKT
New L2 hdr
IP/GRE
PKT


IP Networks
ERPM Source Switch X
Destination Switch Y
Sniffer/Linux Server
Port A
Port B
Port C
Port D
PKT
New L2 hdr
IP/GRE
PKT
New L2 hdr
IP/GRE
PKT

As seen in the above figure, the packets received/transmitted on Port A will be encapsulated with an IP/GRE header plus a new L2 header and sent to the destination ip address (Port D’s ip address) on the sniffer. The Header that gets attached to the packet is 38 bytes long.

If the sniffer does not support IP interface, a destination switch will be needed to receive the encapsulated ERPM packet and locally mirror the whole packet to the Sniffer or a Linux Server.

Decapsulation of ERPM packets at the Destination IP/ Analyzer

- In order to achieve the decapsulation of the original payload from the ERPM header. The below two methods are suggested :

  a. Using Network Analyzer

    - Install any well-known Network Packet Analyzer tool which is open source and free to download.
    - Start capture of ERPM packets on the Sniffer and save it to the trace file (for example : erpmwithheader.pcap).
    - The Header that gets attached to the packet is 38 bytes long. In case of a packet with L3 VLAN, it would be 42 bytes long. The original payload /original mirrored data starts from the

39th byte in a given ERPM packet. The first 38/42 bytes of the header needs to be ignored/ chopped off.

  - Some tools support options to edit the capture file. We can make use of such features (for example: **editcap**) and chop the ERPM header part and save it to a new trace file. This new file (i.e. the original mirrored packet) can be converted back into stream and fed to any egress interface.

b. Using Python script

  - Either have a Linux server's ethernet port ip as the ERPM destination ip or connect the ingress interface of the server to the ERPM MirrorToPort. The analyzer should listen in the forward/ egress interface. If there is only one interface, one can choose the ingress and forward interface to be same and listen in the tx direction of the interface.
  - Download/ Write a small script (for example: **erpm.py**) such that it will strip the given ERPM packet starting from the bit where GRE header ends. Basically all the bits after 0x88BE need to be removed from the packet and sent out through another interface.
  - This script **erpm.zip** is available for download at the following location: http://en.community.dell.com/techcenter/networking/m/force10_networking_scripts/20438882.aspx
  - Unzip the **erpm.zip** and copy the erpm.py file to the Linux server.
  - Run the python script using the following command:

```
python erpm.py -i <ingress interface> -o <egress interface>
```

erpm.py : This is the script downloaded from the script store.

<Ingress interface> : Specify the interface id which is connected to the mirroring port or this should be interface whose ip address has been specified as the destination ip address in the ERPM session.

<Egress interface> : Specify another interface on the Linux server via which the decapsulation packets can Egress. In case there is only one interface, the ingress interface itself can be specified as Egress and the analyzer can listen in the tx direction.

36

Per-VLAN Spanning Tree Plus (PVST+)

Per-VLAN spanning tree plus (PVST+) is supported on the S6000 platform.

Protocol Overview

PVST+ is a variation of spanning tree — developed by a third party — that allows you to configure a separate spanning tree instance for each virtual local area network (VLAN).

For more information about spanning tree, refer to the Spanning Tree Protocol (STP) chapter.

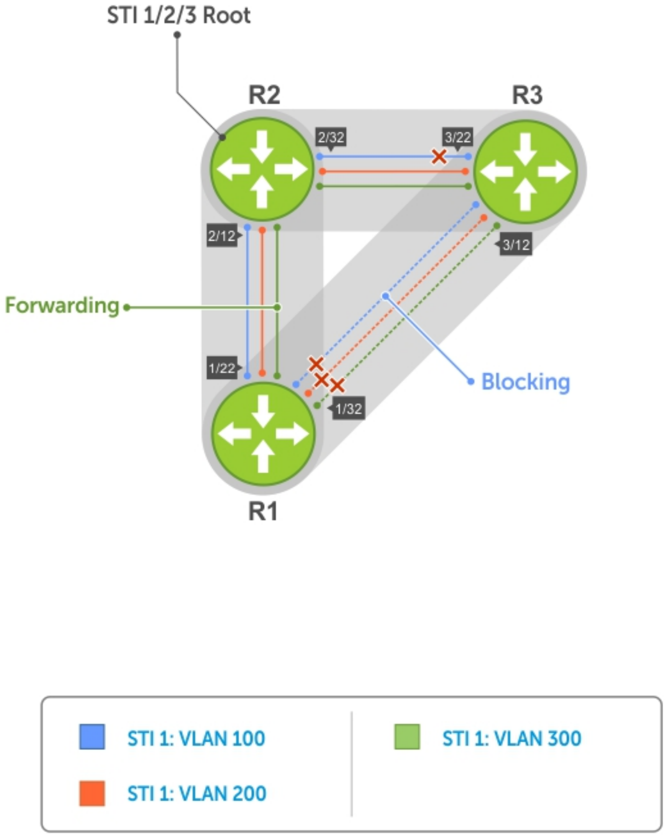

STI 1/2/3 Root
R2
R3
2/32
3/22
2/12
3/12
Forwarding
Blocking
1/22
1/32
R1
STI 1: VLAN 100
STI 1: VLAN 200
STI 1: VLAN 300

Figure 96. Per-VLAN Spanning Tree

The Dell Networking OS supports three other variations of spanning tree, as shown in the following table.

Table 52. Spanning Tree Variations Dell Networking OS Supports

Dell Networking Term	IEEE Specification
Spanning Tree Protocol (STP)	802 .1d
Rapid Spanning Tree Protocol (RSTP)	802 .1w
Multiple Spanning Tree Protocol (MSTP)	802 .1s
Per-VLAN Spanning Tree Plus (PVST+)	Third Party

Implementation Information

- The Dell Networking OS implementation of PVST+ is based on IEEE Standard 802.1w.
- The Dell Networking OS implementation of PVST+ uses IEEE 802.1s costs as the default costs (as shown in the following table). Other implementations use IEEE 802.1w costs as the default costs. If you are using Dell Networking systems in a multivendor network, verify that the costs are values you intended.

Configure Per-VLAN Spanning Tree Plus

Configuring PVST+ is a four-step process.

1. Configure interfaces for Layer 2.
2. Place the interfaces in VLANs.
3. Enable PVST+.
4. Optionally, for load balancing, select a nondefault bridge-priority for a VLAN.

Related Configuration Tasks

- Modifying Global PVST+ Parameters
- Modifying Interface PVST+ Parameters
- Configuring an EdgePort
- Flush MAC Addresses after a Topology Change
- Prevent Network Disruptions with BPDU Guard
- Enabling SNMP Traps for Root Elections and Topology Changes
- Configuring Spanning Trees as Hitless
- PVST+ in Multi-Vendor Networks
- Enabling PVST+ Extended System ID
- PVST+ Sample Configurations

Enabling PVST+

When you enable PVST+, Dell Networking OS instantiates STP on each active VLAN.

1. Enter PVST context.

PROTOCOL PVST mode

```
protocol spanning-tree pvst
```

2. Enable PVST+.

PROTOCOL PVST mode

```
no disable
```

Disabling PVST+

To disable PVST+ globally or on an interface, use the following commands.

- Disable PVST+ globally.
  PROTOCOL PVST mode

  ```
  disable
  ```
- Disable PVST+ on an interface, or remove a PVST+ parameter configuration.
  INTERFACE mode

  ```
  no spanning-tree pvst
  ```

Example of Viewing PVST+ Configuration

To display your PVST+ configuration, use the `show config` command from PROTOCOL PVST mode.

```
Dell_E600(conf-pvst)#show config verbose
!
protocol spanning-tree pvst
  no disable
  vlan 100 bridge-priority 4096
```

Influencing PVST+ Root Selection

As shown in the previous per-VLAN spanning tree illustration, all VLANs use the same forwarding topology because R2 is elected the root, and all GigabitEthernet ports have the same cost.
The following per-VLAN spanning tree illustration changes the bridge priority of each bridge so that a different forwarding topology is generated for each VLAN. This behavior demonstrates how you can use PVST+ to achieve load balancing.

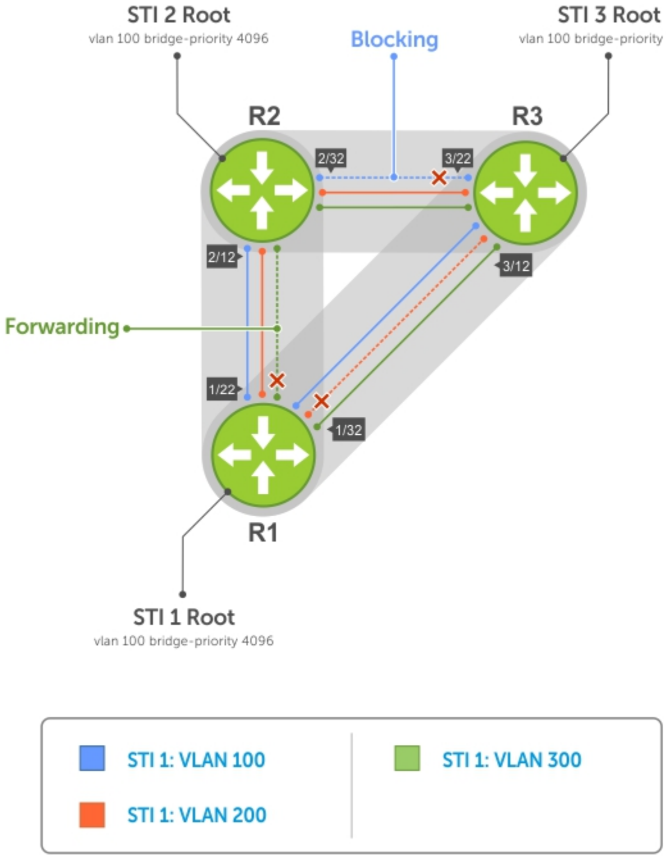

STI 2 Root
vlan 100 bridge-priority 4096
Blocking
STI 3 Root
vlan 100 bridge-priority 4096
R2
R3
2/32
3/22
2/12
3/12
Forwarding
1/22
1/32
R1
STI 1 Root
vlan 100 bridge-priority 4096
STI 1: VLAN 100
STI 1: VLAN 300
STI 1: VLAN 200

Figure 97. Load Balancing with PVST+

The bridge with the bridge value for bridge priority is elected root. Because all bridges use the default priority (until configured otherwise), the lowest MAC address is used as a tie-breaker. To increase the likelihood that a bridge is selected as the STP root, assign bridges a low non-default value for bridge priority.

To assign a bridge priority, use the following command.

- Assign a bridge priority.

  PROTOCOL PVST mode

  `vlan bridge-priority`

  The range is from 0 to 61440.

  The default is **32768**.

Example of the `show spanning-tree pvst vlan` Command

To display the PVST+ forwarding topology, use the `show spanning-tree pvst [vlan vlan-id]` command from EXEC Privilege mode.

```
Dell_E600(conf)#do show spanning-tree pvst vlan 100
VLAN 100
```

```
Root Identifier has priority 4096, Address 0001.e80d.b6d6
Root Bridge hello time 2, max age 20, forward delay 15
Bridge Identifier has priority 4096, Address 0001.e80d.b6d6
Configured hello time 2, max age 20, forward delay 15
We are the root of VLAN 100
Current root has priority 4096, Address 0001.e80d.b6d6
Number of topology changes 5, last change occurred 00:34:37 ago on Gi 1/32

Port 375 (GigabitEthernet 1/22) is designated Forwarding
Port path cost 20000, Port priority 128, Port Identifier 128.375
Designated root has priority 4096, address 0001.e80d.b6:d6
Designated bridge has priority 4096, address 0001.e80d.b6:d6
Designated port id is 128.375 , designated path cost 0
Number of transitions to forwarding state 2
BPDU sent 1159, received 632
The port is not in the Edge port mode

Port 385 (GigabitEthernet 1/32) is designated Forwarding
Port path cost 20000, Port priority 128, Port Identifier 128.385
Designated root has priority 4096, address 0001.e80d.b6:d6
Designated bridge has priority 4096, address 0001.e80d.b6:d6
Designated port id is 128.385 , designated path cost 0
```

Modifying Global PVST+ Parameters

The root bridge sets the values for forward-delay and hello-time, and overwrites the values set on other PVST+ bridges.

- **Forward-delay** — the amount of time an interface waits in the Listening state and the Learning state before it transitions to the Forwarding state.
- **Hello-time** — the time interval in which the bridge sends bridge protocol data units (BPDUs).
- **Max-age** — the length of time the bridge maintains configuration information before it refreshes that information by recomputing the PVST+ topology.

To change PVST+ parameters on the root bridge, use the following commands.

- Change the forward-delay parameter.

  PROTOCOL PVST mode

  ```
  vlan forward-delay
  ```

  The range is from 4 to 30.

  The default is **15 seconds**.
- Change the hello-time parameter.

  PROTOCOL PVST mode

  ```
  vlan hello-time
  ```

  NOTE: With large configurations (especially those configurations with more ports), Dell Networking recommends increasing the hello-time.

  The range is from 1 to 10.

  The default is **2 seconds**.
- Change the max-age parameter.

PROTOCOL PVST mode

```
vlan max-age
```

The range is from 6 to 40.

The default is **20 seconds**.

The values for global PVST+ parameters are given in the output of the `show spanning-tree pvst` command.

Modifying Interface PVST+ Parameters

You can adjust two interface parameters (port cost and port priority) to increase or decrease the probability that a port becomes a forwarding port.

- **Port cost** — a value that is based on the interface type. The greater the port cost, the less likely the port is selected to be a forwarding port.
- **Port priority** — influences the likelihood that a port is selected to be a forwarding port in case that several ports have the same port cost.

The following tables lists the default values for port cost by interface.

Table 53. Default Values for Port Cost

Port Cost	Default Value
100-Mb/s Ethernet interfaces	200000
1-Gigabit Ethernet interfaces	20000
10-Gigabit Ethernet interfaces	2000
Port Channel with 100 Mb/s Ethernet interfaces	180000
Port Channel with 1-Gigabit Ethernet interfaces	18000
Port Channel with 10-Gigabit Ethernet interfaces	1800

NOTE: The Dell Networking OS implementation of PVST+ uses IEEE 802.1s costs as the default costs. Other implementations use IEEE 802.1w costs as the default costs. If you are using Dell Networking systems in a multi-vendor network, verify that the costs are values you intended.

To change the port cost or port priority of an interface, use the following commands.

- Change the port cost of an interface.

  INTERFACE mode

  ```
  spanning-tree pvst vlan cost.
  ```

  The range is from 0 to 200000.

  Refer to the table for the default values.
- Change the port priority of an interface.

  INTERFACE mode

  ```
  spanning-tree pvst vlan priority.
  ```

The range is from 0 to 240, in increments of 16.

The default is **128**.

The values for interface PVST+ parameters are given in the output of the `show spanning-tree pvst` command, as previously shown.

Configuring an EdgePort

The EdgePort feature enables interfaces to begin forwarding traffic approximately 30 seconds sooner. In this mode an interface forwards frames by default until it receives a BPDU that indicates that it should behave otherwise; it does not go through the Learning and Listening states. The `bpduguard shutdown-on-violation` option causes the interface hardware to be shut down when it receives a BPDU. When you only implement `bpduguard`, although the interface is placed in an Error Disabled state when receiving the BPDU, the physical interface remains up and spanning-tree drops packets in the hardware after a BPDU violation. BPDUs are dropped in the software after receiving the BPDU violation.

This feature is the same as PortFast mode in spanning tree.

⚠ **CAUTION: Configure EdgePort only on links connecting to an end station. EdgePort can cause loops if you enable it on an interface connected to a network.**

To enable EdgePort on an interface, use the following command.

- Enable EdgePort on an interface.

  INTERFACE mode

  ```
  spanning-tree pvst edge-port [bpduguard | shutdown-on-violation]
  ```

The EdgePort status of each interface is given in the output of the `show spanning-tree pvst` command, as previously shown.

Dell Networking OS Behavior: Regarding the `bpduguard shutdown-on-violation` command behavior:

- If the interface to be shut down is a port channel, all the member ports are disabled in the hardware.
- When you add a physical port to a port channel already in an Error Disable state, the new member port is also disabled in the hardware.
- When you remove a physical port from a port channel in an Error Disable state, the Error Disabled state is cleared on this physical port (the physical port is enabled in the hardware).
- The `reset linecard` command does not clear the Error Disabled state of the port or the hardware Disabled state. The interface continues to be disables in the hardware.
- You can clear the Error Disabled state with any of the following methods:
  - Perform a `shutdown` command on the interface.
  - Disable the `shutdown-on-violation` command on the interface (the `no spanning-tree stp-id portfast [bpduguard | [shutdown-on-violation]]` command).
  - Disable spanning tree on the interface (the `no spanning-tree` command in INTERFACE mode).
  - Disabling global spanning tree (the `no spanning-tree` command in CONFIGURATION mode).

PVST+ in Multi-Vendor Networks

Some non-Dell Networking systems which have hybrid ports participating in PVST+ transmit two kinds of BPDUs: an 802.1D BPDU and an untagged PVST+ BPDU.

Dell Networking systems do not expect PVST+ BPDU (tagged or untagged) on an untagged port. If this situation occurs, Dell Networking OS places the port in an Error-Disable state. This behavior might result in the network not converging. To prevent Dell Networking OS from executing this action, use the `no spanning-tree pvst err-disable cause invalid-pvst-bpdu` command. After you configure this command, if the port receives a PVST+ BPDU, the BPDU is dropped and the port remains operational.

Enabling PVST+ Extend System ID

In the following example, ports P1 and P2 are untagged members of different VLANs. These ports are untagged because the hub is VLAN unaware. There is no data loop in this scenario; however, you can employ PVST+ to avoid potential misconfigurations.
If you enable PVST+ on the Dell Networking switch in this network, P1 and P2 receive BPDUs from each other. Ordinarily, the Bridge ID in the frame matches the Root ID, a loop is detected, and the rules of convergence require that P2 move to blocking state because it has the lowest port ID.

To keep both ports in a Forwarding state, use extend system ID. Extend system ID augments the bridge ID with a VLAN ID to differentiate BPDUs on each VLAN so that PVST+ does not detect a loop and both ports can remain in a Forwarding state.

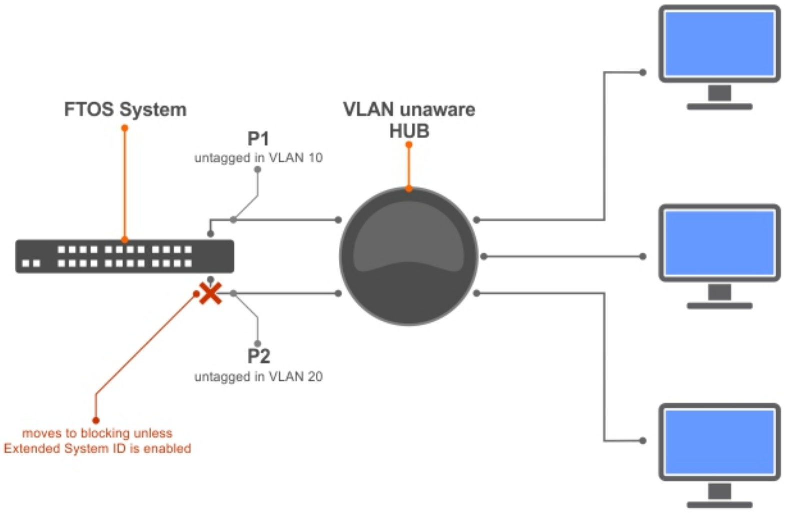


Figure 98. PVST+ with Extend System ID

- Augment the bridge ID with the VLAN ID.
  PROTOCOL PVST mode

  ```
  extend system-id
  ```

Example of Viewing the Extend System ID in a PVST+ Configuration

```
Dell(conf-pvst)#do show spanning-tree pvst vlan 5 brief

VLAN 5
Executing IEEE compatible Spanning Tree Protocol
Root ID Priority 32773, Address 0001.e832.73f7
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID Priority 32773 (priority 32768 sys-id-ext 5), Address 0001.e832.73f7
We are the root of Vlan 5
Configured hello time 2, max age 20, forward delay 15
```

PVST+ Sample Configurations

The following examples provide the running configurations for the topology shown in the previous illustration.

Example of PVST+ Configuration (R1)

```
interface GigabitEthernet 1/22
  no ip address
  switchport
  no shutdown
!
interface GigabitEthernet 1/32
  no ip address
  switchport
  no shutdown
!
protocol spanning-tree pvst
  no disable
  vlan 100 bridge-priority 4096
interface Vlan 100
  no ip address
  tagged GigabitEthernet 1/22,32
  no shutdown
!
interface Vlan 200
  no ip address
  tagged GigabitEthernet 1/22,32
  no shutdown
!
interface Vlan 300
  no ip address
  tagged GigabitEthernet 1/22,32
  no shutdown
!
protocol spanning-tree pvst
  no disable
  vlan 100 bridge-priority 4096
```

Example of PVST+ Configuration (R2)

```
interface GigabitEthernet 2/12
  no ip address
  switchport
  no shutdown
!
interface GigabitEthernet 2/32
  no ip address
  switchport
  no shutdown
!
interface Vlan 100
```

```
  no ip address
  tagged GigabitEthernet 2/12,32
  no shutdown
!
interface Vlan 200
  no ip address
  tagged GigabitEthernet 2/12,32
  no shutdown
!
interface Vlan 300
  no ip address
  tagged GigabitEthernet 2/12,32
  no shutdown
!
protocol spanning-tree pvst
  no disable
  vlan 200 bridge-priority 4096
```

Example of PVST+ Configuration (R3)

```
interface GigabitEthernet 3/12
  no ip address
  switchport
  no shutdown
!
interface GigabitEthernet 3/22
  no ip address
  switchport
  no shutdown
!
interface Vlan 100
  no ip address
  tagged GigabitEthernet 3/12,22
  no shutdown
!
interface Vlan 200
  no ip address
  tagged GigabitEthernet 3/12,22
  no shutdown
!
interface Vlan 300
  no ip address
  tagged GigabitEthernet 3/12,22
  no shutdown
!
protocol spanning-tree pvst
  no disable
  vlan 300 bridge-priority 4096
```

37

Quality of Service (QoS)

Quality of service (QoS) is supported on the S6000 platform.

Differentiated service is accomplished by classifying and queuing traffic, and assigning priorities to those queues.

Table 54. Dell Networking Operating System (OS) Support for Port-Based, Policy-Based Features

Feature	Direction
Port-Based QoS Configurations	Ingress + Egress
Set dot1p Priorities for Incoming Traffic	Ingress
Honor dot1p Priorities on Ingress Traffic	Ingress
Configure Port-based Rate Policing	Ingress
Configure Port-based Rate Shaping	Egress
Policy-Based QoS Configurations	Ingress + Egress
Classify Traffic	Ingress
Create a Layer 3 Class Map	Ingress
Set DSCP Values for Egress Packets Based on Flow	Ingress
Create a Layer 2 Class Map	Ingress
Create a QoS Policy	Ingress + Egress
Create an Input QoS Policy	Ingress
Configure Policy-Based Rate Policing	Ingress
Set a DSCP Value for Egress Packets	Ingress
Set a dot1p Value for Egress Packets	Ingress
Create an Output QoS Policy	Egress
Configure Policy-Based Rate Shaping	Egress
Allocate Bandwidth to the Queue	Egress

Feature	Direction
Configure a Scheduler to Queue	Egress
Specify WRED Drop Precedence	Egress
Create Policy Maps	Ingress + Egress
Create Input Policy Maps	Ingress
Honor DSCP Values on Ingress Packets	Ingress
Honoring dot1p Values on Ingress Packets	Ingress
Create Output Policy Maps	Egress
Specify an Aggregate QoS Policy	Egress
Create Output Policy Maps	Egress
Enabling QoS Rate Adjustment Enabling Strict-Priority Queueing Weighted Random Early Detection	Egress
Create WRED Profiles	Egress

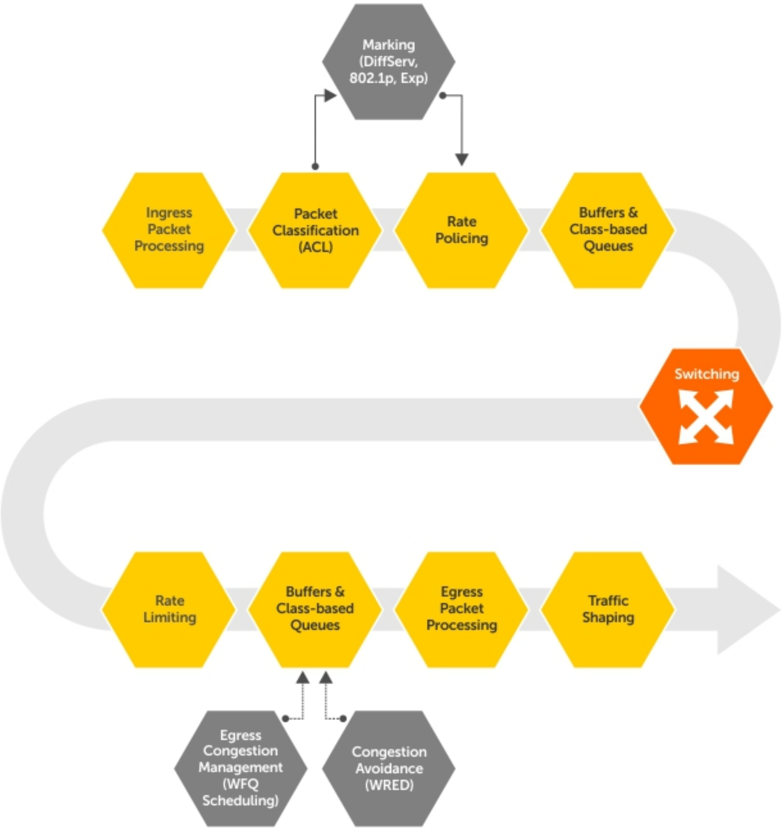

Marking (DiffServ, 802.1p, Exp)
Ingress Packet Processing
Packet Classification (ACL)
Rate Policing
Buffers & Class-based Queues
Switching
Rate Limiting
Buffers & Class-based Queues
Egress Packet Processing
Traffic Shaping
Egress Congestion Management (WFQ Scheduling)
Congestion Avoidance (WRED)

Figure 99. Dell Networking QoS Architecture

Implementation Information

The Dell Networking QoS implementation complies with IEEE 802.1p *User Priority Bits for QoS Indication*.

It also implements these Internet Engineering Task Force (IETF) documents:

- RFC 2474, *Definition of the Differentiated Services Field (DS Field) in the IPv4 Headers*
- RFC 2475, *An Architecture for Differentiated Services*
- RFC 2597, *Assured Forwarding PHB Group*
- RFC 2598, *An Expedited Forwarding PHB*

You cannot configure port-based and policy-based QoS on the same interface.

Port-Based QoS Configurations

You can configure the following QoS features on an interface.

- Setting dot1p Priorities for Incoming Traffic
- Configuring Port-Based Rate Policing
- Configuring Port-Based Rate Shaping

Setting dot1p Priorities for Incoming Traffic

Dell Networking OS places traffic marked with a priority in a queue based on the following table.
If you set a dot1p priority for a port-channel, all port-channel members are configured with the same value. You cannot assign a dot1p value to an individual interface in a port-channel.

Table 55. dot1p-priority Values and Queue Numbers

dot1p	Queue Number
0	2
1	0
2	1
3	3
4	4
5	5
6	6
7	7

- Change the priority of incoming traffic on the interface.

  `dot1p-priority`

Example of Configuring a dot1p Priority on an Interface

```
Dell#config
Dell(conf)#interface tengigabitethernet 1/0
Dell(conf-if)#switchport
Dell(conf-if)#dot1p-priority 1
Dell(conf-if)#end
Dell#
```

Honoring dot1p Priorities on Ingress Traffic

By default, Dell Networking OS does not honor dot1p priorities on ingress traffic.
You can configure this feature on physical interfaces and port-channels, but you cannot configure it on individual interfaces in a port channel.

You can configure service-class dynamic dot1p from CONFIGURATION mode, which applies the configuration to all interfaces. A CONFIGURATION mode service-class dynamic dot1p entry supersedes any INTERFACE entries. For more information, refer to Mapping dot1p Values to Service Queues.

NOTE: You cannot configure `service-policy input` and `service-class dynamic dot1p` on the same interface.

- Honor dot1p priorities on ingress traffic.

  INTERFACE mode

  `service-class dynamic dot1p`

Example of Configuring an Interface to Honor dot1p Priorities on Ingress Traffic

```
Dell#config t
Dell(conf)#interface tengigabitethernet 1/0
Dell(conf-if)#service-class dynamic dot1p
Dell(conf-if)#end
Dell#
```

Priority-Tagged Frames on the Default VLAN

Priority-tagged frames on the default VLAN is available on the S6000 platform.

Priority-tagged frames are 802.1Q tagged frames with VLAN ID 0. For VLAN classification, these packets are treated as untagged. However, the dot1p value is still honored when you configure `service-class dynamic dot1p` or `trust dot1p`.

When priority-tagged frames ingress an untagged port or hybrid port, the frames are classified to the default VLAN of the port and to a queue according to their dot1p priority if you configure `service-class dynamic dotp` or `trust dot1p`. When priority-tagged frames ingress a tagged port, the frames are dropped because, for a tagged port, the default VLAN is 0.

Dell Networking OS Behavior: Hybrid ports can receive untagged, tagged, and priority tagged frames. The rate metering calculation might be inaccurate for untagged ports because an internal assumption is made that all frames are treated as tagged. Internally, the ASIC adds a 4-bytes tag to received untagged frames. Though these 4-bytes are not part of the untagged frame received on the wire, they are included in the rate metering calculation resulting in metering inaccuracy.

Configuring Port-Based Rate Policing

If the interface is a member of a VLAN, you may specify the VLAN for which ingress packets are policed.

- Rate policing ingress traffic on an interface.

  INTERFACE mode

  ```
  rate police
  ```

Example of the `rate police` Command

```
Dell#config t
Dell(conf)#interface tengigabitethernet 1/0
Dell(conf-if)#rate police 100 40 peak 150 50
Dell(conf-if)#end
Dell#
```

Configuring Port-Based Rate Shaping

Configuring port-based rate limiting is supported on the S6000 platform.
Dell Networking OS Behavior: Rate shaping is effectively rate limiting because of its smaller buffer size. Rate shaping on tagged ports is slightly greater than the configured rate and rate shaping on untagged ports is slightly less than configured rate.

Rate shaping buffers, rather than drops, traffic exceeding the specified rate until the buffer is exhausted. If any stream exceeds the configured bandwidth on a continuous basis, it can consume all of the buffer space that is allocated to the port.

- Apply rate shaping to outgoing traffic on a port.

  INTERFACE mode

```
rate shape
```

- Apply rate shaping to a queue.

  QoS Policy mode

  ```
  rate-shape
  ```

Example of `rate shape` Command

```
Dell#config
Dell(conf)#interface tengigabitethernet 1/0
Dell(conf-if)#rate shape 500 50
Dell(conf-if)#end
Dell#
```

Policy-Based QoS Configurations

Policy-based QoS configurations consist of the components shown in the following example.

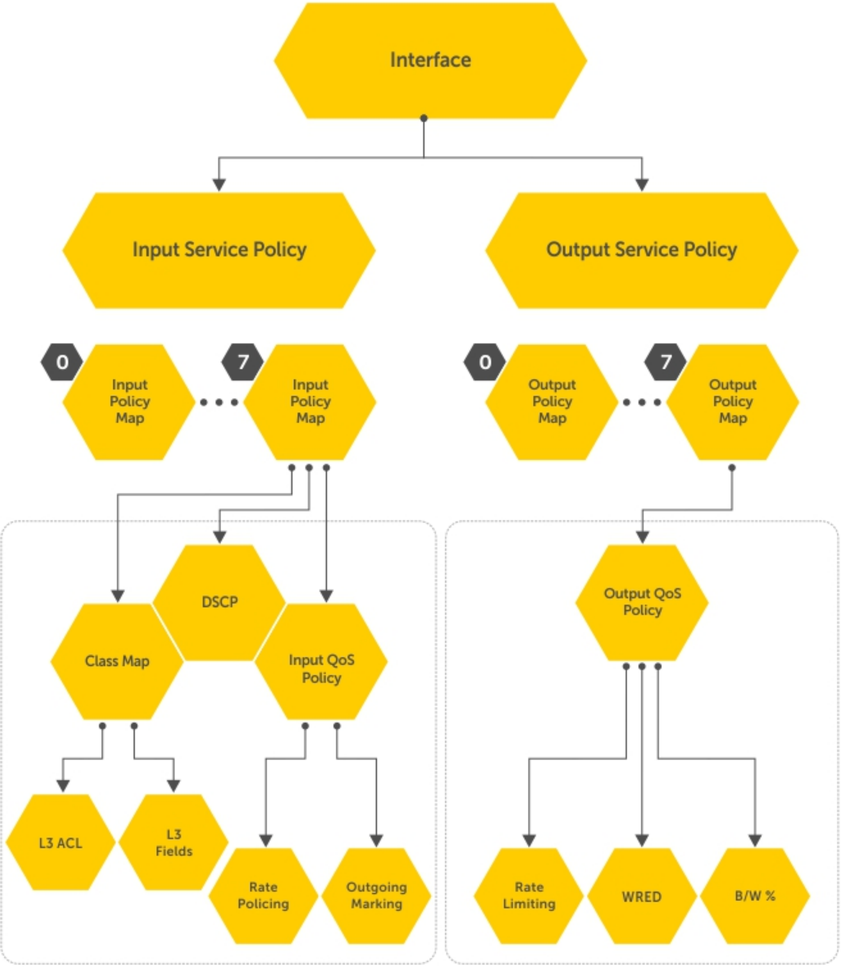

Interface
Input Service Policy
Output Service Policy
0
Input Policy Map
7
Input Policy Map
0
Output Policy Map
7
Output Policy Map
DSCP
Class Map
Input QoS Policy
Output QoS Policy
L3 ACL
L3 Fields
Rate Policing
Outgoing Marking
Rate Limiting
WRED
B/W %

Figure 100. Constructing Policy-Based QoS Configurations

Classify Traffic

Class maps differentiate traffic so that you can apply separate quality of service policies to different types of traffic.

For both class maps, Layer 2 and Layer 3, Dell Networking OS matches packets against match criteria in the order that you configure them.

Creating a Layer 3 Class Map

A Layer 3 class map differentiates ingress packets based on the DSCP value or IP precedence, and characteristics defined in an IP ACL. You can also use VLAN IDs and VRF IDs to classify the traffic using layer 3 class-maps.
You may specify more than one DSCP and IP precedence value, but only one value must match to trigger a positive match for the class map.

NOTE: IPv6 and IP-any class maps cannot match on ACLs or VLANs.

Use step 1 or step 2 to start creating a Layer 3 class map.

1. Create a match-any class map.
   CONFIGURATION mode
   ```
   class-map match-any
   ```
2. Create a match-all class map.
   CONFIGURATION mode
   ```
   class-map match-all
   ```
3. Specify your match criteria.
   CLASS MAP mode
   ```
   match {ip | ipv6 | ip-any}
   ```
   After you create a class-map, Dell Networking OS places you in CLASS MAP mode.

   Match-any class maps allow up to five ACLs. Match-all class-maps allow only one ACL.
4. Link the class-map to a queue.
   POLICY MAP mode
   ```
   service-queue
   ```

Example of Creating a Layer 3 Class Map

```
Dell(conf)#ip access-list standard acl1
Dell(config-std-nacl)#permit 20.0.0.0/8
Dell(config-std-nacl)#exit
Dell(conf)#ip access-list standard acl2
Dell(config-std-nacl)#permit 20.1.1.0/24 order 0
Dell(config-std-nacl)#exit
Dell(conf)#class-map match-all cmap1
Dell(conf-class-map)#match ip access-group acl1
Dell(conf-class-map)#exit
Dell(conf)#class-map match-all cmap2
Dell(conf-class-map)#match ip access-group acl2
Dell(conf-class-map)#exit
```

```
Dell(conf)#policy-map-input pmap
Dell(conf-policy-map-in)#service-queue 3 class-map cmap1
Dell(conf-policy-map-in)#service-queue 1 class-map cmap2
Dell(conf-policy-map-in)#exit
Dell(conf)#interface tegig 1/0
Dell(conf-if-gi-1/0)#service-policy input pmap
```

Examples of Creating a Layer 3 IPv6 Class Map

The following example matches IPv6 traffic with a DSCP value of 40.

```
Dell(conf)# class-map match-all test
Dell(conf-class-map)# match ipv6 dscp 40
```

The following example matches IPv4 and IPv6 traffic with a precedence value of 3.

```
Dell(conf)# class-map match-any test1
Dell(conf-class-map)#match ip-any precedence 3
```

Creating a Layer 2 Class Map

All class maps are Layer 3 by default; however, you can create a Layer 2 class map by specifying the `layer2` option with the `class-map` command.
A Layer 2 class map differentiates traffic according to 802.1p value and/or VLAN and/or characteristics defined in a MAC ACL..

Use Step 1 or Step 2 to start creating a Layer 2 class map.

1. Create a match-any class map.

   CONFIGURATION mode

   ```
   class-map match-any
   ```
2. Create a match-all class map.

   CONFIGURATION mode

   ```
   class-map match-all
   ```
3. Specify your match criteria.

   CLASS MAP mode

   ```
   match mac
   ```

   After you create a class-map, Dell Networking OS places you in CLASS MAP mode.

   Match-any class maps allow up to five access-lists. Match-all class-maps allow only one. You can match against only one VLAN ID.
4. Link the class-map to a queue.

   POLICY MAP mode

   ```
   service-queue
   ```

Determining the Order in Which ACLs are Used to Classify Traffic

When you link class-maps to queues using the `service-queue` command, Dell Networking OS matches the class-maps according to queue priority (queue numbers closer to 0 have lower priorities).
For example, as described in the previous example, class-map *cmap2* is matched against ingress packets before *cmap1*.

ACLs *acl1* and *acl2* have overlapping rules because the address range 20.1.1.0/24 is within 20.0.0.0/8. Therefore (without the keyword `order`), packets within the range 20.1.1.0/24 match positive against *cmap1* and are buffered in queue 7, though you intended for these packets to match positive against *cmap2* and be buffered in queue 4.

In cases such as these, where class-maps with overlapping ACL rules are applied to different queues, use the keyword `order`. Dell Networking OS writes to the CAM ACL rules with lower order numbers (order numbers closer to 0) before rules with higher order numbers so that packets are matched as you intended.

- Specify the order in which you want to apply ACL rules using the keyword `order`.

  ```
  order
  ```

  The order can range from 0 to 254.

  By default, all ACL rules have an order of **255**.

Displaying Configured Class Maps and Match Criteria

To display all class-maps or a specific class map, use the following command.
Dell Networking OS Behavior: An explicit "deny any" rule in a Layer 3 ACL used in a (match any or match all) class-map creates a "default to Queue 0" entry in the CAM, which causes unintended traffic classification. In the following example, traffic is classified in two Queues, 1 and 2. Class-map ClassAF1 is "match any," and ClassAF2 is "match all".

- Display all class-maps or a specific class map.

  EXEC Privilege mode

  ```
  show qos class-map
  ```

Examples of Traffic Classifications

The following example shows incorrect traffic classifications.

```
Dell#show running-config policy-map-input
!
policy-map-input PolicyMapIn
  service-queue 1 class-map ClassAF1 qos-policy QosPolicyIn-1
  service-queue 2 class-map ClassAF2 qos-policy QosPolicyIn-2
Dell#show running-config class-map
!
class-map match-any ClassAF1
  match ip access-group AF1-FB1 set-ip-dscp 10
  match ip access-group AF1-FB2 set-ip-dscp 12
  match ip dscp 10 set-ip-dscp 14
  match ipv6 dscp 20 set-ip-dscp 14
!
class-map match-all ClassAF2
  match ip access-group AF2
  match ip dscp 18

Dell#show running-config ACL
!
ip access-list extended AF1-FB1
  seq 5 permit ip host 23.64.0.2 any
  seq 10 deny ip any any
!
ip access-list extended AF1-FB2
  seq 5 permit ip host 23.64.0.3 any
```

```
  seq 10 deny ip any any
!
ip access-list extended AF2
  seq 5 permit ip host 23.64.0.5 any
  seq 10 deny ip any any

Dell# show cam layer3-qos interface tengigabitethernet 2/49
Cam   Port Dscp Proto Tcp  Src  Dst SrcIp        DstIp     DSCP   Queue
Index                 Flag Port Port                       Marking
-----------------------------------------------------------------------
20416 1   18    IP   0x0    0    0   23.64.0.5/32 0.0.0.0/0 20      2
20417 1   18    IP   0x0    0    0   0.0.0.0/0    0.0.0.0/0 -       0
20418 1   0     IP   0x0    0    0   23.64.0.2/32 0.0.0.0/0 10      1
20419 1   0     IP   0x0    0    0   0.0.0.0/0    0.0.0.0/0 -       0
20420 1   0     IP   0x0    0    0   23.64.0.3/32 0.0.0.0/0 12      1
20421 1   0     IP   0x0    0    0   0.0.0.0/0    0.0.0.0/0 -       0
20422 1   10    0    0x0    0    0   0.0.0.0/0    0.0.0.0/0 14      1
24511 1   0     0    0x0    0    0   0.0.0.0/0    0.0.0.0/0 -       0
```

In the previous example, the ClassAF1 does not classify traffic as intended. Traffic matching the first match criteria is classified to Queue 1, but all other traffic is classified to Queue 0 as a result of CAM entry 20419.

When you remove the explicit “deny any” rule from all three ACLs, the CAM reflects exactly the desired classification.

The following example shows correct traffic classifications.

```
Dell#show cam layer3-qos interface tengigabitethernet 2/49
Cam   Port Dscp Proto Tcp  Src  Dst SrcIp        DstIp     DSCP    Queue
Index                 Flag Port Port                       Marking
-------------------------------------------------------------------------
20416 1     18   IP     0x0  0    0   23.64.0.5/32 0.0.0.0/0 20       2
20417 1     0    IP     0x0  0    0   23.64.0.2/32 0.0.0.0/0 10       1
20418 1     0    IP     0x0  0    0   23.64.0.3/32 0.0.0.0/0 12       1
20419 1     10   0      0x0  0    0   0.0.0.0/0    0.0.0.0/0 14       1
24511 1     0    0      0x0  0    0   0.0.0.0/0    0.0.0.0/0  -       0
```

Dot1p to Queue Mapping Requirement

The dot1p to queue mapping on the system is global and this is used to configure the PRIO2COS table configuration. For DSCP based PFC feature on untagged packets, this mapping must be the same as the default dot1p to queue mapping and should not be changed (as in TABLE 1). If a custom dot1p to queue mapping is present it should be reconfigured to the default dot1p to queue mapping.

- Currently Dell Networking OS supports matching only the following TCP flags:
  - ACK
  - FIN
  - SYN
  - PSH
  - RST
  - URG

In the existing software, ECE/CWR TCP flag qualifiers are not supported.

- Because this functionality forcibly marks all the packets matching the specific match criteria as ‘yellow’, Dell Networking OS does not support Policer based coloring and this feature concurrently.
- If single rate two color policer is configured along with this feature, then by default all packets less than PIR would be considered as “Green” But ‘Green’ packets matching the specific match criteria for which ‘color-marking’ is configured will be over-written and marked as “Yellow”.
- If two rate three color policer is configured along with this feature then,
  - x < CIR – will be marked as “Green”
  - CIR < x< PIR – will be marked as “Yellow”
  - PIR < x – will be marked as “Red”

But ‘Green’ packets matching the specific match criteria for which ‘color-marking’ is configured will be over-written and marked as “Yellow”.

Create a QoS Policy

There are two types of QoS policies — input and output.

Input QoS policies regulate Layer 3 and Layer 2 ingress traffic. The regulation mechanisms for input QoS policies are rate policing and setting priority values.

- **Layer 3** — QoS input policies allow you to rate police and set a DSCP or dot1p value. In addition, you can configure a drop precedence for incoming packets based on their DSCP value by using a DSCP color map. For more information, see DSCP Color Maps.
- **Layer 2** — QoS input policies allow you to rate police and set a dot1p value.

Output QoS policies regulate egress traffic. The regulation mechanisms for output QoS policies are bandwidth percentage, scheduler strict, rate shaping and WRED.

NOTE: When changing a "service-queue" configuration in a QoS policy map, all QoS rules are deleted and re-added automatically to ensure that the order of the rules is maintained. As a result, the Matched Packets value shown in the `show qos statistics` command is reset.

NOTE: To avoid issues misconfiguration causes, Dell Networking recommends configuring either DCBX or Egress QoS features, but not both simultaneously. If you enable both DCBX and Egress QoS at the same time, the DCBX configuration is applied and unexpected behavior occurs on the Egress QoS.

Creating an Input QoS Policy

To create an input QoS policy, use the following steps.

1. Create a Layer 3 input QoS policy.
   CONFIGURATION mode

   ```
   qos-policy-input
   ```

   Create a Layer 2 input QoS policy by specifying the keyword `layer2` after the `qos-policy-input` command.
2. After you create an input QoS policy, do one or more of the following:
   Configuring Policy-Based Rate Policing

   Setting a dot1p Value for Egress Packets

Configuring Policy-Based Rate Policing

To configure policy-based rate policing, use the following command.

- Configure rate police ingress traffic.

  QOS-POLICY-IN mode

  ```
  rate-police
  ```

Setting a dot1p Value for Egress Packets

To set a dot1p value for egress packets, use the following command.

- Set a dscp or dot1p value for egress packets.

  QOS-POLICY-IN mode

  ```
  set mac-dot1p
  ```

Constraints

The systems supporting this feature should use only the default global dot1p to queue mapping configuration as described in Dot1p to Queue Mapping Requirement.

Creating an Output QoS Policy

To create an output QoS policy, use the following commands.

1. Create an output QoS policy.

   CONFIGURATION mode

   ```
   qos-policy-output
   ```
2. After you configure an output QoS policy, do one or more of the following:

   Scheduler Strict — Policy-based Strict-priority Queueing configuration is done through scheduler strict. It is applied to Qos-policy-output. When scheduler strict is applied to multiple Queues, high queue number takes precedence.

   Allocating Bandwidth to Queue

   Specifying WRED Drop Precedence

Configuring Policy-Based Rate Shaping

To configure policy-based rate shaping, use the following command.

- Configure rate shape egress traffic.

  QOS-POLICY-OUT mode

  ```
  rate-shape
  ```

Allocating Bandwidth to Queue

Schedule packets for egress based on Deficit Round Robin (DRR). These strategies both offer a guaranteed data rate.
The following table lists the default bandwidth weights for each queue, and their equivalent percentage which is derived by dividing the bandwidth weight by the sum of all queue weights.

Table 56. Default Bandwidth Weights

Queue	Default Weight	Equivalent Percentage
0	1	6.67%
1	2	13.33%
2	4	26.67%
3	8	53.33%

- Allocate bandwidth to queues.

  `bandwidth-percentage`

Specifying WRED Drop Precedence

Specifying WRED drop precedence is supported on the S6000 platform.

- Specify a WRED profile to yellow and/or green traffic.

  QOS-POLICY-OUT mode

  `wred`

For more information, refer to Applying a WRED Profile to Traffic.

DSCP Color Maps

This section describes how to configure color maps and how to display the color map and color map configuration.

This sections consists of the following topics:

- Creating a DSCP Color Map
- Displaying Color Maps
- Display Color Map Configuration

Creating a DSCP Color Map

You can create a DSCP color map to outline the differentiated services codepoint (DSCP) mappings to the appropriate color mapping (green, yellow, red) for the input traffic. The system uses this information to classify input traffic on an interface based on the DSCP value of each packet and assigns it an initial drop precedence of green, yellow, or red

The default setting for each DSCP value (0-63) is green (low drop precedence). The DSCP color map allows you to set the number of specific DSCP values to yellow or red. Traffic marked as yellow delivers traffic to the egress interface, which will either transmit or drop the packet based on configured queuing behavior. Traffic marked as red (high drop precedence) is dropped.

Important Points to Remember

- All DSCP values that are not specified as yellow or red are colored green (low drop precedence).
- A DSCP value cannot be in both the yellow and red lists. Setting the red or yellow list with any DSCP value that is already in the other list results in an error and no update to that DSCP list is made.

- Each color map can only have one list of DSCP values for each color; any DSCP values previously listed for that color that are not in the new DSCP list are colored green.
- If you configured a DSCP color map on an interface that does not exist or you delete a DSCP color map that is configured on an interface, that interface uses an all green color policy.

To create a DSCP color map:

1. Create the color-aware map QoS DSCP color map.

   CONFIGURATION mode

   `qos dscp-color-map` *`color-map-name`*
2. Create the color aware map profile.

   DSCP-COLOR-MAP

   `dscp {yellow | red} {`*`list-dscp-values`*`}`
3. Apply the map profile to the interface.

   CONFIG-INTERFACE mode

   `qos dscp-color-policy` *`color-map-name`*

Example: Create a DSCP Color Map

The following example creates a DSCP color map profile, color-awareness policy, and applies it to interface **te 0/11**.

Create the DSCP color map profile, **bat-enclave-map**, with a `yellow` drop precedence , and set the DSCP values to `9,10,11,13,15,16`

```
Dell(conf)# qos dscp-color-map bat-enclave-map
Dell(conf-dscp-color-map)# dscp yellow 9,10,11,13,15,16
Dell (conf-dscp-color-map)# exit
```

Assign the color map, **bat-enclave-map** to interface **te 0/11**.

```
Dell(conf)# int te 0/11
Dell(conf-if-te-0/11)# qos dscp-color-policy bat-enclave-map
```

Displaying DSCP Color Maps

To display DSCP color maps, use the **show qos dscp-color-map** command in EXEC mode.

Examples for Creating a DSCP Color Map

Display all DSCP color maps.

```
Dell# show qos dscp-color-map
Dscp-color-map mapONE
  yellow 4,7
  red 20,30
Dscp-color-map mapTWO
  yellow 16,55
```

Display a specific DSCP color map.

```
Dell# show qos dscp-color-map mapTWO
Dscp-color-map mapTWO
  yellow 16,55
```

Displaying a DSCP Color Policy Configuration

To display the DSCP color policy configuration for one or all interfaces, use the `show qos dscp-color-policy {summary [`*`interface`*`] | detail {`*`interface`*`}}` command in EXEC mode.

summary: Displays summary information about a color policy on one or more interfaces.

detail: Displays detailed color policy information on an interface

interface: Enter the name of the interface that has the color policy configured.

Examples for Displaying a DSCP Color Policy

Display summary information about a color policy for one or more interfaces.

```
Dell# show qos dscp-color-policy summary
Interface     dscp-color-map
TE 0/10      mapONE
TE0/11       mapTWO
```

Display summary information about a color policy for a specific interface.

```
Dell# show qos dscp-color-policy summary te 0/10
Interface     dscp-color-map
TE 0/10       mapONE
```

Display detailed information about a color policy for a specific interface

```
Dell# show qos dscp-color-policy detail te 0/10
Interface TenGigabitEthernet 0/10
Dscp-color-map mapONE
  yellow 4,7
  red 20,30
```

Create Policy Maps

There are two types of policy maps: input and output.

Creating Input Policy Maps

There are two types of input policy-maps: Layer 3 and Layer 2.

1. Create a Layer 3 input policy map.
   CONFIGURATION mode

   ```
   policy-map-input
   ```

   Create a Layer 2 input policy map by specifying the keyword `layer2` with the `policy-map-input` command.
2. After you create an input policy map, do one or more of the following:

[Applying a Class-Map or Input QoS Policy to a Queue]()

[Applying an Input QoS Policy to an Input Policy Map]()

[Honoring DSCP Values on Ingress Packets]()

[Honoring dot1p Values on Ingress Packets]()

3. Apply the input policy map to an interface.

Applying a Class-Map or Input QoS Policy to a Queue

To apply a class-map or input QoS policy to a queue, use the following command.

- Assign an input QoS policy to a queue.
  POLICY-MAP-IN mode

  ```
  service-queue
  ```

Applying an Input QoS Policy to an Input Policy Map

To apply an input QoS policy to an input policy map, use the following command.

- Apply an input QoS policy to an input policy map.
  POLICY-MAP-IN mode

  ```
  policy-service-queue <number> qos-polcy
  ```

Honoring DSCP Values on Ingress Packets

Dell Networking OS provides the ability to honor DSCP values on ingress packets using Trust DSCP feature.

The following table lists the standard DSCP definitions and indicates to which queues Dell Networking OS maps DSCP values. When you configure trust DSCP, the matched packets and matched bytes counters are not incremented in the `show qos` statistics.

Table 57. Default DSCP to Queue Mapping

DSCP/CP hex range (XXX)xxx	DSCP Definition	Traditional IP Precedence	Internal Queue ID	DSCP/CP decimal
111XXX		Network Control	3	48–63
110XXX		Internetwork Control	3	48–63
101XXX	EF (Expedited Forwarding)	CRITIC/ECP	2	32–47
100XXX	AF4 (Assured Forwarding)	Flash Override	2	32–47
011XXX	AF3	Flash	1	16–31
010XXX	AF2	Immediate	1	16–31
001XXX	AF1	Priority	0	0–15
000XXX	BE (Best Effort)	Best Effort	0	0–15

Table 58. Default DSCP to Queue Mapping

DSCP/CP hex range (XXX)xxx	DSCP Definition	Traditional IP Precedence	Internal Queue ID	Internal Queue ID (S6000)	DSCP/CP decimal (4–queue)	DSCP/CP decimal (8–queue)
111XXX		Network Control	3	7	48–63	53–63
110XXX		Internetwork Control	3	6	48–63	53–63
101XXX	EF (Expedited Forwarding)	CRITIC/ECP	2	5	32–47	16–23
100XXX	AF4 (Assured Forwarding)	Flash Override	2	4	32–47	16–23
011XXX	AF3	Flash	1	3	16–31	8–15
010XXX	AF2	Immediate	1	2	16–31	8–15
001XXX	AF1	Priority	0	1	0–15	0–7
000XXX	BE (Best Effort)	Best Effort	0	0	0–15	0–7

- Enable the trust DSCP feature.

  POLICY-MAP-IN mode

  ```
  trust diffserv
  ```

Honoring dot1p Values on Ingress Packets

Dell Networking OS honors dot1p values on ingress packets with the Trust dot1p feature.
The following table specifies the queue to which the classified traffic is sent based on the dot1p value.

Table 59. Default dot1p to Queue Mapping

dot1p	Queue ID
0	2
1	0
2	1
3	3
4	4
5	5
6	6
7	7

The dot1p value is also honored for frames on the default VLAN. For more information, refer to Priority-Tagged Frames on the Default VLAN.

- Enable the trust dot1p feature.

  POLICY-MAP-IN mode

  ```
  trust dot1p
  ```

Mapping dot1p Values to Service Queues

All traffic is by default mapped to the same queue, Queue 0.
If you honor dot1p on ingress, you can create service classes based the queueing strategy in Honoring dot1p Values on Ingress Packets. You may apply this queuing strategy globally by entering the following command from CONFIGURATION mode.

- All dot1p traffic is mapped to Queue 0 unless you enable `service-class dynamic dot1p` on an interface or globally.
- Layer 2 or Layer 3 service policies supersede dot1p service classes.

- Create service classes.

  INTERFACE mode

  ```
  service-class dynamic dot1p
  ```

Guaranteeing Bandwidth to dot1p-Based Service Queues

To guarantee bandwidth to dot1p-based service queues, use the following command.
Apply this command in the same way as the `bandwidth-percentage` command in an output QoS policy (refer to Allocating Bandwidth to Queue). The `bandwidth-percentage` command in QOS-POLICY-OUT mode supersedes the `service-class bandwidth-percentage` command.

- Guarantee a minimum bandwidth to queues globally.

  CONFIGURATION mode

  ```
  service-class bandwidth-percentage
  ```

Applying an Input Policy Map to an Interface

To apply an input policy map to an interface, use the following command.
You can apply the same policy map to multiple interfaces, and you can modify a policy map after you apply it.

- You cannot apply an input Layer 2 QoS policy on an interface you also configure with vlan-stack access.
- If you apply a service policy that contains an ACL to more than one interface, Dell Networking OS uses ACL optimization to conserve CAM space. The ACL optimization behavior detects when an ACL exists in the CAM rather than writing it to the CAM multiple times.

- Apply an input policy map to an interface.

  INTERFACE mode

  ```
  service-policy input
  ```

  Specify the keyword `layer2` if the policy map you are applying a Layer 2 policy map.

Creating Output Policy Maps

Creating output policy maps is supported on the S6000 platform.

1. Create an output policy map.

   CONFIGURATION mode

   ```
   policy-map-output
   ```
2. After you create an output policy map, do one or more of the following:

Applying an Output QoS Policy to a Queue

Specifying an Aggregate QoS Policy

Applying an Output Policy Map to an Interface

3. Apply the policy map to an interface.

Applying an Output QoS Policy to a Queue

To apply an output QoS policy to a queue, use the following command.

- Apply an output QoS policy to queues.
  INTERFACE mode

  ```
  service-queue
  ```

Specifying an Aggregate QoS Policy

To specify an aggregate QoS policy, use the following command.

- Specify an aggregate QoS policy.
  POLICY-MAP-OUT mode

  ```
  policy-aggregate
  ```

Applying an Output Policy Map to an Interface

To apply an output policy map to an interface, use the following command.

- Apply an input policy map to an interface.
  INTERFACE mode

  ```
  service-policy output
  ```

  You can apply the same policy map to multiple interfaces, and you can modify a policy map after you apply it.

Enabling QoS Rate Adjustment

By default while rate limiting, policing, and shaping, Dell Networking OS does not include the Preamble, SFD, or the IFG fields. These fields are overhead; only the fields from MAC destination address to the CRC are used for forwarding and are included in these rate metering calculations.
The Ethernet packet format consists of:

- Preamble: 7 bytes Preamble
- Start frame delimiter (SFD): 1 byte
- Destination MAC address: 6 bytes
- Source MAC address: 6 bytes
- Ethernet Type/Length: 2 bytes
- Payload: (variable)
- Cyclic redundancy check (CRC): 4 bytes
- Inter-frame gap (IFG): (variable)

You can optionally include overhead fields in rate metering calculations by enabling QoS rate adjustment.

QoS rate adjustment is disabled by default.

- Specify the number of bytes of packet overhead to include in rate limiting, policing, and shaping calculations.
  CONFIGURATION mode

  ```
  qos-rate-adjust overhead-bytes
  ```

  For example, to include the Preamble and SFD, type `qos-rate-adjust 8`. For variable length overhead fields, know the number of bytes you want to include.

  The default is disabled.

Enabling Strict-Priority Queueing

Strict-priority means that Dell Networking OS de-queues all packets from the assigned queue before servicing any other queues.

- The `strict-priority` supersedes `bandwidth-percentage` configuration.
- A queue with strict priority can starve other queues in the same port-pipe.

- Assign strict priority to one unicast queue.
  CONFIGURATION mode

  ```
  strict-priority
  ```

  The range is from 1 to 3.

Queue Classification Requirements for PFC Functionality

Queue classification requirements for PFC functionality are mentioned below:

- On untagged ports, Queue classification must be based on DSCP.
- On tagged ports, Queue classification must be based on Dot1p. Layer 3 classification configurations should not be present on the port.
- On hybrid ports, Queue classification can be based on either Dot1p (for tagged packets) or DSCP (for untagged packets) but not both.

Example Case:

PFC does not work for tagged traffic, when DSCP based class map is applied on a hybrid port or on a tagged port. Assume two switches A and B are connected back to back.

Consider the case where untagged packets arrive on switch A, if you want to generate PFC for priority 2 for DSCP range 0-7, then you have to match the interested traffic. You must use the class map and associate to queue 1 using the policy map. The same class map needs to be applied in switch B as well and when queue 1 gets congested, PFC would be generated for priority 2. Switch A on receiving PFC frames with priority 2 would stop scheduling queue 1.

If a tagged packet with VLAN dot1p as 5 ingresses on switch A. Consider that tagged packet also has DSCP in range of 0-7.These packets will match the class map and get queued on queue 1 on both the switches.

But when queue 1 gets congested on switch B, PFC frames for tagged packets will not be generated as PFC is not enabled on dot1p priority 5.

Support for marking dot1p value in L3 Input Qos Policy

In case the incoming packet is untagged and the packet which goes out to the peer is tagged, then the dot1p should be marked appropriately using L3 Input Qos Policy. This is required because in the peer switch PFC will be generated based on the dot1p value. Currently if the ingress is untagged and egress is tagged, then dot1p priority 0(default) will be added as part of the tag header and from the next hop PFC will be based on that dot1p priority. Support is added to mark the dot1p value in the L3 Input Qos Policy in this feature. Hence it is possible to mark both DSCP and Dot1p simultaneously in the L3 Input Qos Policy. You are expected to mark the Dot1p priority when the ingress packets are untagged but go out to the peer as tagged

NOTE: L2 qos-policy behavior will be retained and would not be changed, that is we would not allow to set both DSCP and Dot1p in the L2 Input Qos Policy.

Example case:

Consider that two switches A and B are connected back to back via a tagged interface. Consider the case where untagged packets arrive on switch A, if you want to generate PFC for priority 2 for DSCP range 0-7, then you must need to match the interested traffic using the class map. You should create an L3 Input Qos Policy and mark vlan dot1p as 2. You have to associate both the L3 class map and L3 Input Qos Policy to queue 1 using the policy map.

In switch B, global dot1p honoring should be enabled, this will queue the packets on queue 1 as the dot1p will be 2 and PFC should be enabled for priority 2. The policy map applied on switch A need not be enabled in switch B. When queue 1 in switch B gets congested, PFC will be generated for priority 2 which will be honored in switch A.

You will not get the below CLI errors after adding this support:

```
Dell(conf)#qos-policy-input qos-input
Dell(conf-qos-policy-in)#set mac-dot1p 5
```

% Error: Dot1p marking is not allowed on L3 Input Qos Policy.

```
Dell(conf-qos-policy-in)#
```

You will also be able to mark both DSCP and Dot1p in the L3 Input Qos Policy:

```
Dell(conf)#qos-policy-input qos-input
Dell(conf-qos-policy-in)#set mac-dot1p 2
Dell(conf-qos-policy-in)#set ip-dscp 5
Dell Dell(conf-qos-policy-in)#
```

Weighted Random Early Detection

Weighted random early detection (WRED) is supported on the S6000 platform.

The WRED congestion avoidance mechanism drops packets to prevent buffering resources from being consumed.

Traffic is a mixture of various kinds of packets. The rate at which some types of packets arrive might be greater than others. In this case, the space on the buffer and traffic manager (BTM) (ingress or egress) can be consumed by only one or a few types of traffic, leaving no space for other types. You can apply a

WRED profile to a policy-map so that specified traffic can be prevented from consuming too much of the BTM resources.

WRED uses a profile to specify minimum and maximum threshold values. The minimum threshold is the allotted buffer space for specified traffic, for example, 1000KB on egress. If the 1000KB is consumed, packets are dropped randomly at an exponential rate until the maximum threshold is reached (as shown in the following illustration); this procedure is the "early detection" part of WRED. If the maximum threshold, for example, 2000KB, is reached, all incoming packets are dropped until the buffer space consumes less than 2000KB of the specified traffic.

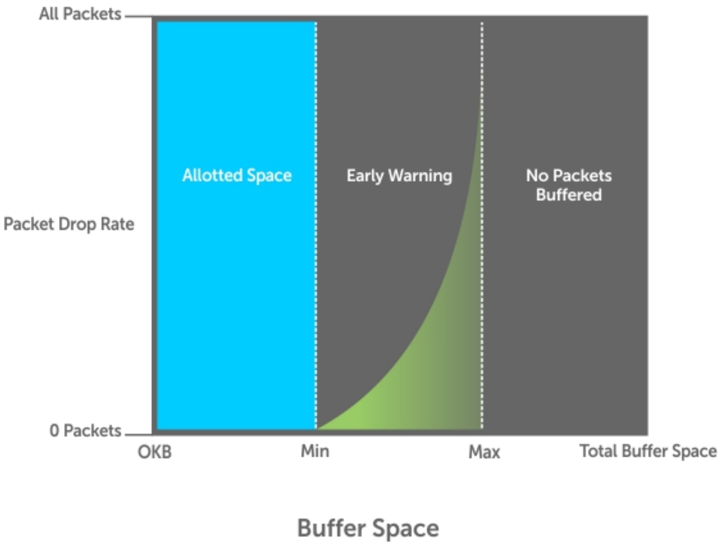


Figure 101. Packet Drop Rate for WRED

You can create a custom WRED profile or use one of the five pre-defined profiles.

Creating WRED Profiles

To create WRED profiles, use the following commands.

1. Create a WRED profile.
   CONFIGURATION mode

   ```
   wred-profile
   ```
2. Specify the minimum and maximum threshold values.
   WRED mode

   ```
   threshold
   ```

Applying a WRED Profile to Traffic

After you create a WRED profile, you must specify to which traffic Dell Networking OS should apply the profile.
Dell Networking OS assigns a color (also called drop precedence) — red, yellow, or green — to each packet based on it DSCP value before queuing it.

DSCP is a 6–bit field. Dell Networking uses the first three bits (LSB) of this field (DP) to determine the drop precedence.

- DP values of 110 and 100, 101 map to yellow; all other values map to green.
- If you do not configure Dell Networking OS to honor DSCP values on ingress (refer to Honoring DSCP Values on Ingress Packets), all traffic defaults to green drop precedence.

- Assign a WRED profile to either yellow or green traffic.

  QOS-POLICY-OUT mode

  ```
  wred
  ```

Displaying Default and Configured WRED Profiles

To display the default and configured WRED profiles, use the following command.

- Display default and configured WRED profiles and their threshold values.

  EXEC mode

  ```
  show qos wred-profile
  ```

Displaying WRED Drop Statistics

To display WRED drop statistics, use the following command.

- Display the number of packets Dell Networking OS the WRED profile drops.

  EXEC Privilege mode

  ```
  show qos statistics wred-profile
  ```

Pre-Calculating Available QoS CAM Space

Pre-calculating available QoS CAM space is supported on the S6000 platform.
Before Dell Networking OS version 7.3.1, there was no way to measure the number of CAM entries a policy-map would consume (the number of CAM entries that a rule uses is not predictable; from 1 to 16 entries might be used per rule depending upon its complexity). Therefore, it was possible to apply to an interface a policy-map that requires more entries than are available. In this case, the system writes as many entries as possible, and then generates an CAM-full error message (shown in the following example). The partial policy-map configuration might cause unintentional system behavior.

```
     %EX2YD:12 %DIFFSERV-2-DSA_QOS_CAM_INSTALL_FAILED: Not enough space in L3
Cam(PolicyQos) for class 2 (TeGi 12/20) entries on portpipe 1
```

The `test cam-usage` command allows you to verify that there are enough available CAM entries before applying a policy-map to an interface so that you avoid exceeding the QoS CAM space and partial configurations. This command measures the size of the specified policy-map and compares it to the available CAM space in a partition for a specified port-pipe.

Test the policy-map size against the CAM space for a specific port-pipe or all port-pipes using these commands:

- `test cam-usage service-policy input` *`policy-map`* `{stack-unit }` *`number`* `port-set` *`number`*
- `test cam-usage service-policy input` *`policy-map`* `{stack-unit }` *`all`*

The output of this command, shown in the following example, displays:

- The estimated number of CAM entries the policy-map will consume.
- Whether or not the policy-map can be applied.
- The number of interfaces in a port-pipe to which the policy-map can be applied.

Specifically:

- **Available CAM** — the available number of CAM entries in the specified CAM partition for the specified line card or stack-unit port-pipe.
- **Estimated CAM** — the estimated number of CAM entries that the policy will consume when it is applied to an interface.
- **Status** — indicates whether the specified policy-map can be completely applied to an interface in the port-pipe.
  - **Allowed** — indicates that the policy-map can be applied because the estimated number of CAM entries is less or equal to the available number of CAM entries. The number of interfaces in the port-pipe to which the policy-map can be applied is given in parentheses.
  - **Exception** — indicates that the number of CAM entries required to write the policy-map to the CAM is greater than the number of available CAM entries, and therefore the policy-map cannot be applied to an interface in the specified port-pipe.

NOTE: The `show cam-usage` command provides much of the same information as the `test cam-usage` command, but whether a policy-map can be successfully applied to an interface cannot be determined without first measuring how many CAM entries the policy-map would consume; the `test cam-usage` command is useful because it provides this measurement.

- Verify that there are enough available CAM entries.

  `test cam-usage`

Example of the `test cam-usage` Command

```
Dell# test cam-usage service-policy input pmap_l2  port-set 0

Port-pipe | CAM Partition | Available CAM | Estimated CAM | Status
===================================================================
  0           L2ACL           500             200         Allowed(2)
```

Specifying Policy-Based Rate Shaping in Packets Per Second

Configuration of rate shaping in packets per second (pps) for QoS output policies is supported on the S6000 platform.

You can configure rate shaping in pps for a QoS output policy, apart from specifying the rate shaping value in bytes. You can also configure the peak rate and the committed rate for packets in kilobits per second (Kbps) or pps.

Committed rate refers to the guaranteed bandwidth for traffic entering or leaving the interface under normal network conditions. When traffic propagates at an average rate that is less than or equal to the committed rate, it is considered to be green-colored or coded. When the transmitted traffic falls below the committed rate, the bandwidth, which is not used by any traffic that is traversing the network, is aggregated to form the committed burst size. Traffic is considered to be green-colored up to the point at which the unused bandwidth does not exceed the committed burst size.

Peak rate refers to the maximum rate for traffic arriving or exiting an interface under normal traffic conditions. Peak burst size indicates the maximum size of unused peak bandwidth that is aggregated. This aggregated bandwidth enables brief durations of burst traffic that exceeds the peak rate and committed burst.

In releases of Dell Networking OS earlier than Release 9.3(0.0), you can configure only the maximum shaping attributes, such as the peak rate and the peak burst settings. You can now specify the committed or minimum burst and committed rate attributes. The committed burst and committed rate values can be defined either in bytes or pps.

You can use the `rate-shape pps peak-rate burst-packets` command in the QoS Policy Out Configuration mode to configure the peak rate and burst size as a measure of pps. Alternatively, you can use the `rate shape kbps peak-rate burst-KB` command to configure the peak rate and peak burst size as a measure of bytes.

Similarly, you can use the `rate-shape pps peak-rate burst-packets committed pps committed-rate burst-packets` command in the QoS Policy Out Configuration mode to configure the committed rate and committed burst size as a measure of pps. Alternatively, you can use the `rate shape kbps peak-rate burst-KB committed kbps committed-rate burst-KB` command to configure the committed rate and committed burst size as a measure of bytes. If you configure the peak rate in pps, the peak burst size must also be configured as a measure of number of packets. Similarly, if you configure the peak rate in Kbps, the peak burst size must also be configured as a measure of bytes.

Configuring Policy-Based Rate Shaping

Configuration of rate shaping for QoS output policies in packets per second (pps) is supported on the S6000 platform.

You can explicitly specify the rate shaping functionality for QoS output policies as peak rate and committed rate attributes. You can also configure the peak burst and committed burst sizes. All of these settings can be configured in Kbps, Mbps, or pps.

To configure the peak and committed rates and burst sizes, perform the following steps:

1. Configure the peak rate and peak burst size in pps in QoS Policy Out Configuration mode.

   QOS-POLICY-OUT mode

   ```
   Dell(config-qos-policy-out)# rate shape pps peak-rate burst-packets
   ```

2. Alternatively, configure the peak rate and peak burst size in bytes.

   QOS-POLICY-OUT mode

   ```
   Dell(config-qos-policy-out)# rate shape Kbps peak-rate burst-KB
   ```

3. Configure the committed rate and committed burst size in pps.

   QOS-POLICY-OUT mode

```
Dell(config-qos-policy-out)# rate shape pps peak-rate burst-packets
committed pps committed-rate burst-packets
```

4. Alternatively, configure the committed rate and committed burst size in bytes.

   QOS-POLICY-OUT mode

   ```
   Dell(config-qos-policy-out)# rate shape Kbps peak-rate burst-KB committed
   Kbps committed-rate burst-KB
   ```

Configuring Weights and ECN for WRED

The feature to configure a weight factor for weighted random early detection (WRED) and Explicit Congestion Notification (ECN) functionality for backplane ports is supported on the Additionally, the feature to configure a weight for WRED and ECN functionality for front-end ports is supported on the S6000 platform.

The WRED congestion avoidance functionality drops packets to prevent buffering resources from being consumed. Traffic is a mixture of various kinds of packets. The rate at which some types of packets arrive might be greater than others. In this case, the space on the buffer and traffic manager (BTM) (ingress or egress) can be consumed by only one or few types of traffic, leaving no space for other types. You can apply a WRED profile to a policy-map so that the specified traffic can be prevented from consuming too much of the BTM resources.

WRED drops packets when the average queue length exceeds the configured threshold value to signify congestion. ECN is a capability that enhances WRED by marking the packets instead of causing WRED to drop them when the threshold value is exceeded. If you configure ECN for WRED, devices employ ECN to mark the packets and reduce the rate of sending packets in a congested network.

In a best-effort network topology, data packets are transmitted in a manner in which latency or throughput is not maintained to be at an effective level. Packets are dropped when the network experiences a large traffic load. This best-effort network deployment is not suitable for applications that are time-sensitive, such as video on demand (VoD) or voice over IP (VoIP) applications. In such cases, you can use ECN in conjunction with WRED to resolve the dropping of packets under congested conditions.

Using ECN, the packets are marked for transmission at a later time after the network recovers from the heavy traffic state to an optimal load. In this manner, enhanced performance and throughput are achieved. Also, the devices can respond to congestion before a queue overflows and packets are dropped, enabling improved queue management.

When a packet reaches the device with ECN enabled for WRED, the average queue size is computed. To measure the average queue size, a weight factor is used. This weight factor is user-configurable. You can use the `wred weight number` command to configure the weight for the WRED average queue size. The mark probability value is the number of packets dropped when the average queue size reaches the maximum threshold value.

The weight factor is set to zero by default, which causes the same behavior as dropping of packets by WRED during network loads or also called instantaneous ECN marking. In a topology in which congestion of the network varies over time, you can specify a weight to enable a smooth, seamless averaging of packets to handle the sudden overload of packets based on the previous time sampling performed. You can specify the weight parameter for front-end and backplane ports separately in the range of 0 through 15.

You can enable WRED and ECN capabilities per queue for granularity. You can disable these functionality per queue, and you can also specify the minimum and maximum buffer thresholds for each color-coding of the packets. You can configure maximum drop rate percentage of yellow and green profiles. You can set up these parameters for both front-end and backplane ports.

Global Service Pools With WRED and ECN Settings

A global buffer pool, whichis a shared buffer pool accessed by multiple queues when the minimum guaranteed buffers for the queue are consumed, can be configured on the S6000 platform.

Support for global service pools is now available. You can configure global service pools that are shared buffer pools accessed by multiple queues when the minimum guaranteed buffers for the queue are consumed. S6000 platform support four global service-pools in the egress direction. Two service pools are used– one for loss-based queues and the other for lossless (priority-based flow control (PFC)) queues. You can enable WRED and ECN configuration on the global service-pools.

You can define WRED profiles and weight on each of the global service-pools for both loss-based and lossless (PFC) service- pools. The following events occur when you configure WRED and ECN on global service-pools:

- If WRED/ECN is enabled on the global service-pool with threshold values and if it is not enabled on the queues, WRED/ECN are not effective based on global service-pool WRED thresholds. The queue on which the traffic is scheduled must contain WRED/ECN settings, which are enabled for WRED, to be valid for that traffic.
- When WRED is configured on the global service-pool (regardless of whether ECN on global service-pool is configured), and one or more queues have WRED enabled and ECN disabled, WRED is effective for the minimum of the thresholds between the queue threshold and the service-pool threshold.
- When WRED is configured on the global service-pool (regardless of whether ECN on global service-pool is configured), and one or more queues are enabled with both WRED and ECN, ECN marking takes effect. The packets are ECN marked up to shared- buffer limits as determined by the shared-ratio for that global service-pool.

WRED/ECN configurations for the queues that belong to backplane ports are common to all the backplane ports and cannot be specified separately for each backplane port granularity. This behavior occurs to prevent system-level complexities in enabling this support for backplane ports. Also, WRED/ECN is not supported for multicast packets.

The following table describes the WRED and ECN operations that occur for various scenarios of WRED and ECN configuration on the queue and service pool. (X denotes not-applicable in the table, 1 indicates that the setting is enabled, 0 represents a disabled setting.)

Table 60. Scenarios of WRED and ECN Configuration

Queue Configuration		Service-Pool Configuration		WRED Threshold Relationship Q threshold = Q-T, Service pool threshold = SP-T	Expected Functionality
WRED	ECN	WRED	ECN		
0	0	X	X	X	WRED/ECN not applicable
1	0	0	X	X	Queue based WRED, No ECN marking
		1	X	Q-T < SP-T	
				SP-T < Q-T	SP based WRED, No ECN marking
1	1	0	X	X	Queue-based ECN marking above queue threshold.
		1	X	Q-T < SP-T	ECN marking to shared buffer limits of the service-pool and then packets are tail dropped.
				SP-T < Q-T	Same as above but ECN marking starts above SP-T.

Configuring WRED and ECN Attributes

The functionality to configure a weight factor for the WRED and ECN functionality for backplane ports is supported on the Z9000 platform. Additionally, the functionality to configure a weight for WRED and ECN functionality for front-end ports is supported on the S6000 platform.A global buffer pool that is a shared buffer pool accessed by multiple queues when the minimum guaranteed buffers for the queue are consumed can be configured on the S6000 platform.

WRED drops packets when the average queue length exceeds the configured threshold value to signify congestion. Explicit Congestion Notification (ECN) is a capability that enhances WRED by marking the packets instead of causing WRED to drop them when the threshold value is exceeded. If you configure ECN for WRED, devices employ this functionality of ECN to mark the packets and reduce the rate of sending packets in a congested, heavily-loaded network.

To configure the weight factor for WRED and ECN capabilities, global buffer pools for multiple queues, and associating a service class with ECN marking, perform the following:

1. Configure the weight factor for the computation of average-queue size. This weight value applies to front-end ports.

   QOS-POLICY-OUT mode

   ```
   Dell(conf-qos-policy-out)#wred—profile weight number
   ```
2. Configure a WRED profile, and specify the threshold and maximum drop rate.

   WRED mode

```
Dell(conf-wred) #wred—profile thresh-1

Dell(conf-wred) #threshold min 100 max 200 max-drop-rate 40
```

3. Configure another WRED profile, and specify the threshold and maximum drop rate.

   WRED mode

   ```
   Dell(conf-wred) #wred—profile thresh-2

   Dell(conf-wred) #threshold min 300 max 400 max-drop-rate 80
   ```

4. Associate the service class with the WRED profile, and assign the WRED profile to specific queues on backplane ports.

   CONFIGURATION mode

   ```
   Dell(conf) #service-class wred green queue5 thresh-1 queue7 thresh-2
   backplane

   Dell(conf) #service-class wred yellow queue1 thresh-2 queue3 thresh-1
   backplane

   Dell(conf) #service-class wred weight queue0 11 queue6 4 queue7 9 backplane
   ```

5. Create a global buffer pool that is a shared buffer pool accessed by multiple queues when the minimum guaranteed buffers for the queue are consumed. S6000 platform supports four global service-pools in the egress direction.

   mode

   ```
   Dell(conf) #service-pool wred green pool0 thresh-1 pool1 thresh-2
   Dell(conf) #service-pool wred yellow pool0 thresh-3 pool1 thresh-4

   Dell(conf) #service-pool wred weight pool0 11 pool1 4
   ```

6. Attach the ECN marking to specific queues on backplane ports with a service class.

   CONFIGURATION mode

   ```
   Dell(conf) #service-class wred ecn 0,3-5,7 backplane
   ```

7. Create a service class and associate the threshold weight of the shared buffer with each of the queues per port in the egress direction.

   INTERFACE mode

   ```
   Dell(conf-if-te-0/8)#Service-class buffer shared-threshold-weight
   ```

Guidelines for Configuring ECN for Classifying and Color-Marking Packets

Keep the following points in mind while configuring the marking and mapping of incoming packets using ECN fields in IPv4 headers:

- Currently Dell Networking OS supports matching only the following TCP flags:
  - ACK

  - FIN
  - SYN
  - PSH
  - RST
  - URG

In the existing software, ECE/CWR TCP flag qualifiers are not supported.

- Because this functionality forcibly marks all the packets matching the specific match criteria as 'yellow', Dell Networking OS does not support Policer based coloring and this feature concurrently.
- If single rate two color policer is configured along with this feature, then by default all packets less than PIR would be considered as "Green" But 'Green' packets matching the specific match criteria for which 'color-marking' is configured will be over-written and marked as "Yellow".
- If two rate three color policer is configured along with this feature then,
  - x < CIR – will be marked as "Green"
  - CIR < x< PIR – will be marked as "Yellow"
  - PIR < x – will be marked as "Red"

But 'Green' packets matching the specific match criteria for which 'color-marking' is configured will be over-written and marked as "Yellow".

Sample configuration to mark non-ecn packets as "yellow" with Multiple traffic class

Consider the example where there are no different traffic classes that is all the packets are egressing on the default 'queue0'.
Dell Networking OS can be configured as below to mark the non-ecn packets as yellow packets.

```
!
ip access-list standard ecn_0
 seq 5 permit any ecn 0

class-map match-any ecn_0_cmap
 match ip access-group ecn_0 set-color  yellow

!
policy-map-input ecn_0_pmap
 service-queue 0 class-map ecn_0_cmap
```

Applying this policy-map "ecn_0_pmap" will mark all the packets with 'ecn == 0' as yellow packets on queue0 (default queue).

Classifying Incoming Packets Using ECN and Color-Marking

Explicit Congestion Notification (ECN) is a capability that enhances WRED by marking the packets instead of causing WRED to drop them when the threshold value is exceeded. If you configure ECN for WRED, devices employ this functionality of ECN to mark the packets and reduce the rate of sending packets in a congested, heavily-loaded network. This feature is supported in S6000 platform.

ECN is a mechanism using which network switches indicate congestion to end hosts for initiating appropriate action. End hosts uses two least significant bits of ToS to indicate that it is ECT. When intermediate network node encounters congestion, remarks ECT to CE for end host to take appropriate action. During congestion, ECN enabled packets are not subject to any kind of drops like WRED except tail drops. Though ECN & WRED are independent technologies, BRCM has made WRED a mandatory for ECN to work.

On ECN deployment, the non-ECN packets that are transmitted on the ECN-WRED enabled interface will be considered as Green packets and will be subject to the early WRED drops. Typically the TCP-acks, OAM, ICMP ping packets will be non-ECN in nature and it is not desirable for this packets getting WRED dropped.

In such a condition, it is necessary that the switch is capable to take differentiated actions for ECN/Non-ECN packets. After classifying packets to ECN/Non-ECN, marking ECN and Non-ECN packets to different color packets is performed.

Policy based ingress QOS involves the following three steps to achieve QOS:

1. Classification of incoming traffic.
2. Specify the differentiated actions for different traffic class.
3. Attach the policy-map to the interface.

Dell Networking OS support different types of match qualifiers to classify the incoming traffic.

Match qualifiers can be directly configured in the class-map command or it can be specified through one or more ACL which in turn specifies the combination of match qualifiers.

Until Release 9.3(0.0), support is available for classifying traffic based on the 6-bit DSCP field of the IPv4 packet.

As a part of this feature, the 2-bit ECN field of the IPv4 packet will also be available to be configured as one of the match qualifier. This way the entire 8-bit ToS field of the IPv4 header shall be used to classify traffic.

The Dell Networking OS Release 9.3(0.0) supports the following QOS actions in the ingress policy based QOS:

1. Rate Policing
2. Queuing
3. Marking

For the L3 Routed packets, the DSCP marking is the only marking action supported in the software. As a part of this feature, the additional marking action to set the "color" of the traffic will be provided.

Until Release 9.3(0.0), the software has the capability to qualify only on the 6-bit DSCP part of the ToS field in IPv4 Header. You can now accept and process incoming packets based on the 2-bit ECN part of the ToS field in addition to the DSCP categorization. The IPv4 ACLs (standard and Extended) are enhanced to add this qualifier. This new keyword 'ecn' is present for all L3 ACL types (TCP/UDP/IP/ICMP) at the level where the 'DSCP' qualifier is positioned in the current ACL commands.

Dell Networking OS supports the capability to contain DSCP and ECN classifiers simultaneously for the same ACL entry.

You can use the ecn keyword with the ip access-list standard, ip access-list extended, seq, and permit commands for standard and extended IPv4 ACLs to match incoming packets with the specified ECN values.

Similar to 'dscp' qualifier in the existing L3 ACL command, the 'ecn' qualifier can be used along with all other supported ACL match qualifiers such as SIP/DIP/TCP/UDP/SRC PORT/DST PORT/ ICMP.

Until Release 9.3(0.0), ACL supports classification based on the below TCP flags:

- ACK
- FIN
- SYN
- PSH
- RST
- URG

You can now use the 'ecn' match qualifier along with the above TCP flag for classification.

The following combination of match qualifiers is acceptable to be configured for the Dell Networking OS software through L3 ACL command:

- Classification based on DSCP only
- Classification based on ECN only
- Classification based on ECN and DSCP concurrently

You can now use the set-color yellow keyword with the match ip access-group command to mark the color of the traffic as 'yellow' would be added in the 'match ip' sequence of the class-map configuration.

By default, all packets are considered as 'green' (without the rate-policer and trust-diffserve configuration) and hence support would be provided to mark the packets as 'yellow' alone will be provided.

By default Dell Networking OS drops all the 'RED' or 'violate' packets.

The following combination of marking actions to be specified match sequence of the class-map command:

- set a new DSCP for the packet
- set the packet color as 'yellow'
- set the packet color as 'yellow' and set a new DSCP for the packet

This marking action to set the color of the packet is allowed only on the 'match-any' logical operator of the class-map.

This marking-action can be configured for all of the below L3 match sequence types:

- match ip access-group
- match ip dscp
- match ip precedence

- match ip vlan

Sample configuration to mark non-ecn packets as “yellow” with single traffic class

Consider the use case where the packet with DSCP value “40” need to be enqueued in queue#2 and packets with DSCP value as 50 need to be enqueued in queue#3. And all the packets with ecn value as ‘0’ must be marked as ‘yellow’.
The above requirement can be achieved using either of the two approaches.

The above requirement can be achieved using either of the two approaches.

Approach without explicit ECN match qualifiers for ECN packets:

```
!
ip access-list standard dscp_50
 seq 5 permit any dscp 50
!
ip access-list standard dscp_40
 seq 5 permit any dscp 40
!
ip access-list standard dscp_50_non_ecn
 seq 5 permit any dscp 50 ecn 0

!
ip access-list standard dscp_40_non_ecn
 seq 5 permit any dscp 40 ecn 0
!
class-map match-any class_dscp_40
 match ip access-group dscp_40_non_ecn set-color yellow
 match ip access-group dscp_40
!
class-map match-any class_dscp_50
 match ip access-group dscp_50_non_ecn set-color yellow
 match ip access-group dscp_50
!
policy-map-input pmap_dscp_40_50
 service-queue 2 class-map class_dscp_40
 service-queue 3 class-map class_dscp_50
```

Approach with explicit ECN match qualifiers for ECN packets:

```
!
ip access-list standard dscp_50_ecn
 seq 5 permit any dscp 50 ecn 1
 seq 10 permit any dscp 50 ecn 2
 seq 15 permit any dscp 50 ecn 3

!
ip access-list standard dscp_40_ecn
 seq 5 permit any dscp 40 ecn 1
 seq 10 permit any dscp 40 ecn 2
 seq 15 permit any dscp 40 ecn 3

!
ip access-list standard dscp_50_non_ecn
 seq 5 permit any dscp 50 ecn 0

!
ip access-list standard dscp_40_non_ecn
 seq 5 permit any dscp 40 ecn 0
!
```

```
class-map match-any class_dscp_40
 match ip access-group dscp_40_non_ecn set-color yellow
 match ip access-group dscp_40_ecn
!
class-map match-any class_dscp_50
 match ip access-group dscp_50_non_ecn set-color yellow
 match ip access-group dscp_50_ecn
!
policy-map-input pmap_dscp_40_50
 service-queue 2 class-map class_dscp_40
 service-queue 3 class-map class_dscp_50
```

Applying Layer 2 Match Criteria on a Layer 3 Interface

To process Layer 3 packets that contain a dot1p (IEEE 802.1p) VLAN Layer 2 header, configure VLAN tags on a Layer 3 port interface which is configured with an IP address but has no VLAN associated with it. You can also configure a VLAN sub-interface on the port interface and apply a policy map that classifies packets using the dot1p VLAN ID.

To apply an input policy map with Layer 2 match criteria to a Layer 3 port interface, use the `service-policy input policy-name layer 2` command in Interface configuration mode.

To apply a Layer 2 policy on a Layer 3 interface:

1. Configure an interface with an IP address or a VLAN sub-interface

   CONFIGURATION mode

   ```
   Dell(conf)# interface fo 0/0
   ```

   INTERFACE mode

   ```
   Dell(conf-if-fo-0/0)# ip address 90.1.1.1/16
   ```
2. Configure a Layer 2 QoS policy with Layer 2 (Dot1p or source MAC-based) match criteria.

   CONFIGURATION mode

   ```
   Dell(conf)# policy-map-input l2p layer2
   ```
3. Apply the Layer 2 policy on a Layer 3 interface.

   INTERFACE mode

   ```
   Dell(conf-if-fo-0/0)# service-policy input l2p layer2
   ```

Managing Hardware Buffer Statistics

Bufffer statistics tracking utility is supported on the S6000 platform.

The memory management unit (MMU) on S6000 platform is 12.2 MB in size. It contains approximately 60,000 cells, each of which is 208 bytes in size. MMU also has another portion of 3 MB allocated to it. The entire MMU space is shared across a maximum of 104 logical ports to support the egress admission-control functionality to implement scheduling and shaping on per-port and per-queue levels. Also, the MMU buffer cells can be used by each port or queue either as a static partition or as a dynamic partition. With dynamic mode, you can specify the percentage of available buffer that is utilized by a queue. This dynamic partition or block is set to be two-thirds of the available buffers for all unicast queues and one-fifth of the available buffers for all multicast queues on these platforms.

The maximum number of ports, including fan-out, supported is 104 and the maximum number of queues supported is 21. Analyzing and evaluating buffer statistics enables monitoring of resources and tuning of allocation of buffers. Max Use count mode provides the maximum values of counters accumulated over a period of time. Current Use count mode enables you to obtain a snapshot of the counters, at a particular time, using a triggering utility. The trigger can either be software-based or based on a predetermined threshold event. Software-based triggers are supported, which are the values derived from the show command output in the Max Use count mode. In Dell Networking OS Release 9.3(0.0), only the Max Use count mode of operation is supported for the computation of maximum counter values.

Depending on the buffer space statistical values that you can obtain, you can modify the settings for buffer area to achieve enhanced reliability and efficiency in the handling of packets. This evaluation and administration of buffer statistics is useful and important in deployments that experience congestion frequently. The receive buffer must be large enough to save all data that is received when the system processes a PFC PAUSE frame.

You can use the `service-class buffer shared-threshold-weight queue0 ... queue7` *`number`* command in Interface Configuration mode to specify the threshold weight for the shared buffer for each of the queues per port.

1. Create a 10-Gigabit Ethernet interface.

   ```
   Dell(conf)#interface TenGigabitEthernet 0/8
   ```

2. Configure the threshold weight of the shared buffer for the queues you want. In this example, this setting is configured for queues 5 and 7.

   ```
   Dell(conf-if-te-0/8)#Service-class buffer shared-threshold-weight queue5 4
   queue7 6
   ```

Enabling Buffer Statistics Tracking

This functionality is supported on the S6000 platform.

You can enable the tracking of statistical values of buffer spaces at a global level. The buffer statistics tracking utility operates in the max use count mode that enables the collection of maximum values of counters.

To configure the buffer statistics tracking utility, perform the following step:

Enable the buffer statistics tracking utility and enter the Buffer Statistics Snapshot configuration mode.
CONFIGURATION mode

```
Dell(conf)#buffer-stats-snapshot
```

```
Dell(conf)#no disable
```

You must enable this utility to be able to configure the parameters for buffer statistics tracking. By default, buffer statistics tracking is disabled.

38

Routing Information Protocol (RIP)

Routing information protocol (RIP) is supported on the S6000 platform.

RIP is based on a distance-vector algorithm; it tracks distances or hop counts to nearby routers when establishing network connections.

RIP protocol standards are listed in the Standards Compliance chapter.

Protocol Overview

RIP is the oldest interior gateway protocol.

There are two versions of RIP: RIP version 1 (RIPv1) and RIP version 2 (RIPv2). These versions are documented in RFCs 1058 and 2453.

RIPv1

RIPv1 learns where nodes in a network are located by automatically constructing a routing data table.

The routing table is established after RIP sends out one or more broadcast signals to all adjacent nodes in a network. Hop counts of these signals are tracked and entered into the routing table, which defines where nodes in the network are located.

The information that is used to update the routing table is sent as either a request or response message. In RIPv1, automatic updates to the routing table are performed as either one-time requests or periodic responses (every 30 seconds). RIP transports its responses or requests by means of user datagram protocol (UDP) over port 520.

RIP must receive regular routing updates to maintain a correct routing table. Response messages containing a router’s full routing table are transmitted every 30 seconds. If a router does not send an update within a certain amount of time, the hop count to that route is changed to unreachable (a route hop metric of 16 hops). Another timer sets the amount of time before the unreachable routes are removed from the routing table.

This first RIP version does not support variable length subnet mask (VLSM) or classless inter-domain routing (CIDR) and is not widely used.

RIPv2

RIPv2 adds support for subnet fields in the RIP routing updates, thus qualifying it as a classless routing protocol.

The RIPv2 message format includes entries for route tags, subnet masks, and next hop addresses. Another enhancement included in RIPv2 is multicasting for route updates on IP multicast address 224.0.0.9.

Implementation Information

Dell Networking OS supports both versions of RIP and allows you to configure one version globally and the other version on interfaces or both versions on the interfaces.

The following table lists the defaults for RIP in Dell Networking OS.

Table 61. RIP Defaults

Feature	Default
Interfaces running RIP	• Listen to RIPv1 and RIPv2 • Transmit RIPv1
RIP timers	• update timer = 30 seconds • invalid timer = 180 seconds • holddown timer = 180 seconds • flush timer = 240 seconds
Auto summarization	Enabled
ECMP paths supported	16

Configuration Information

By default, RIP is disabled in Dell Networking OS.

To configure RIP, you must use commands in two modes: ROUTER RIP and INTERFACE. Commands executed in the ROUTER RIP mode configure RIP globally, while commands executed in the INTERFACE mode configure RIP features on that interface only.

RIP is best suited for small, homogeneous networks. You must configure all devices within the RIP network to support RIP if they are to participate in the RIP.

Configuration Task List

The following is the configuration task list for RIP.

- Enabling RIP Globally (mandatory)
- Configure RIP on Interfaces (optional)
- Controlling RIP Routing Updates (optional)
- Setting Send and Receive Version (optional)
- Generating a Default Route (optional)
- Controlling Route Metrics (optional)
- Summarize Routes (optional)
- Controlling Route Metrics
- Debugging RIP

For a complete listing of all commands related to RIP, refer to the *Dell Networking OS Command Reference Interface Guide*.

Enabling RIP Globally

By default, RIP is not enabled in Dell Networking OS.
To enable RIP globally, use the following commands.

1. Enter ROUTER RIP mode and enable the RIP process on Dell Networking OS.
   CONFIGURATION mode

   ```
   router rip
   ```
2. Assign an IP network address as a RIP network to exchange routing information.
   ROUTER RIP mode

   ```
   network ip-address
   ```

Examples of Verifying RIP is Enabled and Viewing RIP Routes

After designating networks with which the system is to exchange RIP information, ensure that all devices on that network are configured to exchange RIP information.

The Dell Networking OS default is to send RIPv1 and to receive RIPv1 and RIPv2. To change the RIP version globally, use the `version` command in ROUTER RIP mode.

To view the global RIP configuration, use the `show running-config` command in EXEC mode or the `show config` command in ROUTER RIP mode.

```
Dell(conf-router_rip)#show config
!
router rip
  network 10.0.0.0
Dell(conf-router_rip)#
```

When the RIP process has learned the RIP routes, use the `show ip rip database` command in EXEC mode to view those routes.

```
Dell#show ip rip database
Total number of routes in RIP database: 978
160.160.0.0/16
    [120/1] via 29.10.10.12, 00:00:26, Fa 0/0
160.160.0.0/16    auto-summary
2.0.0.0/8
    [120/1] via 29.10.10.12, 00:01:22, Fa 0/0
2.0.0.0/8         auto-summary
4.0.0.0/8
    [120/1] via 29.10.10.12, 00:01:22, Fa 0/0
4.0.0.0/8         auto-summary
8.0.0.0/8
    [120/1] via 29.10.10.12, 00:00:26, Fa 0/0
8.0.0.0/8         auto-summary
12.0.0.0/8
    [120/1] via 29.10.10.12, 00:00:26, Fa 0/0
12.0.0.0/8        auto-summary
20.0.0.0/8
    [120/1] via 29.10.10.12, 00:00:26, Fa 0/0
20.0.0.0/8        auto-summary
29.10.10.0/24 directly connected,Fa 0/0
29.0.0.0/8        auto-summary
31.0.0.0/8
    [120/1] via 29.10.10.12, 00:00:26, Fa 0/0
31.0.0.0/8        auto-summary
```

```
192.162.2.0/24
    [120/1] via 29.10.10.12, 00:01:21, Fa 0/0
192.162.2.0/24    auto-summary
192.161.1.0/24
    [120/1] via 29.10.10.12, 00:00:27, Fa 0/0
192.161.1.0/24    auto-summary
192.162.3.0/24
    [120/1] via 29.10.10.12, 00:01:22, Fa 0/0
192.162.3.0/24    auto-summary
```

To disable RIP globally, use the `no router rip` command in CONFIGURATION mode.

Configure RIP on Interfaces

When you enable RIP globally on the system, interfaces meeting certain conditions start receiving RIP routes.

By default, interfaces that you enable and configure with an IP address in the same subnet as the RIP network address receive RIPv1 and RIPv2 routes and send RIPv1 routes.

Assign IP addresses to interfaces that are part of the same subnet as the RIP network identified in the `network` command syntax.

Controlling RIP Routing Updates

By default, RIP broadcasts routing information out all enabled interfaces, but you can configure RIP to send or to block RIP routing information, either from a specific IP address or a specific interface.
To control which devices or interfaces receive routing updates, configure a direct update to one router and configure interfaces to block RIP updates from other sources.

To control the source of RIP route information, use the following commands.

- Define a specific router to exchange RIP information between it and the Dell Networking system.

  ROUTER RIP mode

  ```
  neighbor ip-address
  ```

  You can use this command multiple times to exchange RIP information with as many RIP networks as you want.
- Disable a specific interface from sending or receiving RIP routing information.

  ROUTER RIP mode

  ```
  passive-interface interface
  ```

Assigning a Prefix List to RIP Routes

Another method of controlling RIP (or any routing protocol) routing information is to filter the information through a prefix list. A prefix list is applied to incoming or outgoing routes.
Those routes must meet the conditions of the prefix list; if not, Dell Networking OS drops the route. Prefix lists are globally applied on all interfaces running RIP. Configure the prefix list in PREFIX LIST mode prior to assigning it to the RIP process.

For configuration information about prefix lists, refer to Access Control Lists (ACLs).

To apply prefix lists to incoming or outgoing RIP routes, use the following commands.

- Assign a configured prefix list to all incoming RIP routes.

  ROUTER RIP mode

```
distribute-list prefix-list-name in
```

- Assign a configured prefix list to all outgoing RIP routes.

  ROUTER RIP mode

  ```
  distribute-list prefix-list-name out
  ```

To view the current RIP configuration, use the `show running-config` command in EXEC mode or the `show config` command in ROUTER RIP mode.

Adding RIP Routes from Other Instances

In addition to filtering routes, you can add routes from other routing instances or protocols to the RIP process.
With the `redistribute` command, you can include open shortest path first (OSPF), static, or directly connected routes in the RIP process.

To add routes from other routing instances or protocols, use the following commands.

- Include directly connected or user-configured (static) routes in RIP.

  ROUTER RIP mode

  ```
  redistribute {connected | static} [metric metric-value] [route-map map-name]
  ```

  – *metric-value*: the range is from 0 to 16.

  – *map-name*: the name of a configured route map.
- Include specific OSPF routes in RIP.

  ROUTER RIP mode

  ```
  redistribute ospf process-id [match external {1 | 2} | match internal]
  [metric value] [route-map map-name]
  ```

  Configure the following parameters:

  – *process-id*: the range is from 1 to 65535.

  – `metric`: the range is from 0 to 16.

  – *map-name*: the name of a configured route map.

To view the current RIP configuration, use the `show running-config` command in EXEC mode or the `show config` command in ROUTER RIP mode.

Setting the Send and Receive Version

To change the RIP version globally or on an interface in Dell Networking OS, use the following command.
To specify the RIP version, use the `version` command in ROUTER RIP mode. To set an interface to receive only one or the other version, use the `ip rip send version` or the `ip rip receive version` commands in INTERFACE mode.

You can set one RIP version globally on the system using `system`. This command sets the RIP version for RIP traffic on the interfaces participating in RIP unless the interface was specifically configured for a specific RIP version.

- Set the RIP version sent and received on the system.

  ROUTER RIP mode

  ```
  version {1 | 2}
  ```

- Set the RIP versions received on that interface.
  INTERFACE mode

  ```
  ip rip receive version [1] [2]
  ```
- Set the RIP versions sent out on that interface.
  INTERFACE mode

  ```
  ip rip send version [1] [2]
  ```

Examples of the RIP Process

To see whether the `version` command is configured, use the `show config` command in ROUTER RIP mode.

The following example shows the RIP configuration after the `ROUTER RIP mode version` command is set to RIPv2. When you set the `ROUTER RIP mode version` command, the interface (GigabitEthernet 0/0) participating in the RIP process is also set to send and receive RIPv2 (shown in bold).

To view the routing protocols configuration, use the `show ip protocols` command in EXEC mode.

```
Dell#show ip protocols

  Routing Protocols is RIP
  Sending updates every 30 seconds, next due in 23
  Invalid after 180 seconds, hold down 180, flushed after 240
  Output delay 8 milliseconds between packets
  Automatic network summarization is in effect
  Outgoing filter for all interfaces is
  Incoming filter for all interfaces is
  Default redistribution metric is 1
  Default version control: receive version 2, send version 2
       Interface        Recv  Send
       GigabitEthernet 0/0  2   2
  Routing for Networks:
       10.0.0.0

  Routing Information Sources:
  Gateway      Distance          Last Update

  Distance: (default is 120)

Dell#
```

To configure an interface to receive or send both versions of RIP, include 1 and 2 in the command syntax. The command syntax for sending both RIPv1 and RIPv2 and receiving only RIPv2 is shown in the following example.

```
Dell(conf-if)#ip rip send version 1 2
Dell(conf-if)#ip rip receive version 2
```

The following example of the `show ip protocols` command confirms that both versions are sent out that interface. This interface no longer sends and receives the same RIP versions as Dell Networking OS does globally (shown in bold).

```
Dell#show ip protocols
  Routing Protocols is RIP
  Sending updates every 30 seconds, next due in 11
  Invalid after 180 seconds, hold down 180, flushed after 240
  Output delay 8 milliseconds between packets
```

```
  Automatic network summarization is in effect
  Outgoing filter for all interfaces is
  Incoming filter for all interfaces is
  Default redistribution metric is 1
  Default version control: receive version 2, send version 2
         Interface          Recv   Send
         FastEthernet 0/0  2    1  2
  Routing for Networks:
         10.0.0.0

Routing Information Sources:
  Gateway      Distance        Last Update

  Distance: (default is 120)

Dell#
```

Generating a Default Route

Traffic is forwarded to the default route when the traffic's network is not explicitly listed in the routing table.
Default routes are not enabled in RIP unless specified. Use the `default-information originate` command in ROUTER RIP mode to generate a default route into RIP. In Dell Networking OS, default routes received in RIP updates from other routes are advertised if you configure the `default-information originate` command.

- Specify the generation of a default route in RIP.

  ROUTER RIP mode

  ```
  default-information originate [always] [metric value] [route-map route-map-name]
  ```

  - `always`: Enter the keyword `always` to always generate a default route.
  - *`value`* The range is from 1 to 16.
  - *`route-map-name`*: The name of a configured route map.

To confirm that the default route configuration is completed, use the `show config` command in ROUTER RIP mode.

Summarize Routes

Routes in the RIPv2 routing table are summarized by default, thus reducing the size of the routing table and improving routing efficiency in large networks.

By default, the `autosummary` command in ROUTER RIP mode is enabled and summarizes RIP routes up to the classful network boundary.

If you must perform routing between discontiguous subnets, disable automatic summarization. With automatic route summarization disabled, subnets are advertised.

The `autosummary` command requires no other configuration commands. To disable automatic route summarization, enter `no autosummary` in ROUTER RIP mode.

NOTE: If you enable the `ip split-horizon` command on an interface, the system does not advertise the summarized address.

Controlling Route Metrics

As a distance-vector protocol, RIP uses hop counts to determine the best route, but sometimes the shortest hop count is a route over the lowest-speed link.
To manipulate RIP routes so that the routing protocol prefers a different route, manipulate the route by using the `offset` command.

Exercise caution when applying an `offset` command to routers on a broadcast network, as the router using the `offset` command is modifying RIP advertisements before sending out those advertisements.

The `distance` command also allows you to manipulate route metrics. To assign different weights to routes so that the ones with the lower weight or administrative distance assigned are preferred, use the `distance` command.

To set route matrixes, use the following commands.

- Apply a weight to all routes or a specific route and ACL.

  ROUTER RIP mode

  ```
  distance weight [ip-address mask [access-list-name]]
  ```

  Configure the following parameters:

  - *`weight`*: the range is from 1 to 255. The default is **120**.
  - *`ip-address mask`*: the IP address in dotted decimal format (A.B.C.D), and the mask in slash format (/x).
  - *`access-list-name`*: the name of a configured IP ACL.
- Apply an additional number to the incoming or outgoing route metrics.

  ROUTER RIP mode

  ```
  offset-list access-list-name {in | out} offset [interface]
  ```

  Configure the following parameters:

  - *`prefix-list-name`*: the name of an established Prefix list to determine which incoming routes are modified
  - *`offset`*: the range is from 0 to 16.
  - *`interface`*: the type, slot, and number of an interface.

To view the configuration changes, use the `show config` command in ROUTER RIP mode.

Debugging RIP

The `debug ip rip` command enables RIP debugging.
When you enable debugging, you can view information on RIP protocol changes or RIP routes.

To enable RIP debugging, use the following command.

- `debug ip rip [interface | database | events | trigger]`

  EXEC privilege mode

  Enable debugging of RIP.

Example of the `debug ip rip` Command

The following example shows the confirmation when you enable the debug function.

```
Dell#debug ip rip
RIP protocol debug is ON
Dell#
```

To disable RIP, use the `no debug ip rip` command.

RIP Configuration Example

The examples in this section show the command sequence to configure RIPv2 on the two routers shown in the following illustration — *Core 2* and *Core 3*.

The host prompts used in the following example reflect those names. The examples are divided into the following groups of command sequences:

- Configuring RIPv2 on Core 2
- Core 2 RIP Output
- RIP Configuration on Core 3
- Core 3 RIP Output
- RIP Configuration Summary

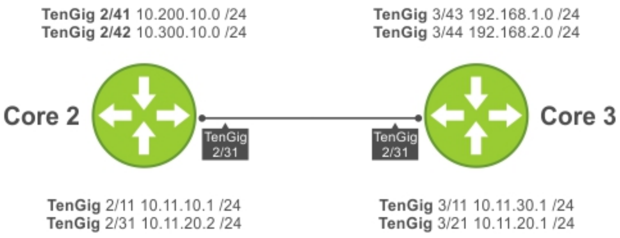

TenGig 2/41 10.200.10.0 /24
TenGig 2/42 10.300.10.0 /24
TenGig 3/43 192.168.1.0 /24
TenGig 3/44 192.168.2.0 /24
Core 2
TenGig 2/31
TenGig 2/31
Core 3
TenGig 2/11 10.11.10.1 /24
TenGig 2/31 10.11.20.2 /24
TenGig 3/11 10.11.30.1 /24
TenGig 3/21 10.11.20.1 /24

Figure 102. RIP Topology Example

RIP Configuration on Core2

The following example shows how to configure RIPv2 on a host named Core2.

Example of Configuring RIPv2 on Core 2

```
Core2(conf-if-gi-2/31)#
Core2(conf-if-gi-2/31)#router rip
Core2(conf-router_rip)#ver 2
Core2(conf-router_rip)#network 10.200.10.0
Core2(conf-router_rip)#network 10.300.10.0
Core2(conf-router_rip)#network 10.11.10.0
Core2(conf-router_rip)#network 10.11.20.0
Core2(conf-router_rip)#show config
!
router rip
  network 10.0.0.0
  version 2
Core2(conf-router_rip)#
```

Core 2 RIP Output

The examples in the section show the core 2 RIP output.

Examples of the `show ip` Commands to View Core 2 Information

- To display Core 2 RIP database, use the `show ip rip database` command.
- To display Core 2 RIP setup, use the `show ip route` command.
- To display Core 2 RIP activity, use the `show ip protocols` command.

The following example shows the `show ip rip database` command to view the learned RIP routes on Core 2.

```
Core2(conf-router_rip)#end
00:12:24: %RPM0-P:CP %SYS-5-CONFIG_I: Configured from console by console
Core2#show ip rip database
Total number of routes in RIP database: 7
10.11.30.0/24
    [120/1] via 10.11.20.1, 00:00:03, GigabitEthernet 2/31
10.300.10.0/24     directly connected,GigabitEthernet 2/42
10.200.10.0/24     directly connected,GigabitEthernet 2/41
10.11.20.0/24      directly connected,GigabitEthernet 2/31
10.11.10.0/24      directly connected,GigabitEthernet 2/11
10.0.0.0/8         auto-summary
192.168.1.0/24
    [120/1] via 10.11.20.1, 00:00:03, GigabitEthernet 2/31
192.168.1.0/24     auto-summary
192.168.2.0/24
    [120/1] via 10.11.20.1, 00:00:03, GigabitEthernet 2/31
192.168.2.0/24     auto-summary
Core2#
```

The following example shows the `show ip route` command to show the RIP setup on Core 2.

```
Core2#show ip route

Codes: C - connected, S - static, R - RIP,
       B - BGP, IN - internal BGP, EX - external BGP,LO - Locally Originated,
       O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
       N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
       E2 - OSPF external type 2, i - IS-IS, L1 - IS-IS level-1,
       L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
       > - non-active route, + - summary route

Gateway of last resort is not set

Destination Gateway            Dist/Metric Last Change
----------- ------- ----------- -----------
C    10.11.10.0/24   Direct, Gi 2/11          0/0   00:02:26
C    10.11.20.0/24   Direct, Gi 2/31          0/0   00:02:02
R    10.11.30.0/24   via 10.11.20.1, Gi 2/31  120/1 00:01:20
C    10.200.10.0/24  Direct, Gi 2/41          0/0   00:03:03
C    10.300.10.0/24  Direct, Gi 2/42          0/0   00:02:42
R    192.168.1.0/24  via 10.11.20.1, Gi 2/31  120/1 00:01:20
R    192.168.2.0/24  via 10.11.20.1, Gi 2/31  120/1 00:01:20
Core2#
R    192.168.1.0/24  via 10.11.20.1, Gi 2/31  120/1 00:05:22
R    192.168.2.0/24  via 10.11.20.1, Gi 2/31  120/1 00:05:22

Core2#
```

The following example shows the `show ip protocols` command to show the RIP configuration activity on Core 2.

```
Core2#show ip protocols
Routing Protocol is "RIP"
  Sending updates every 30 seconds, next due in 17
  Invalid after 180 seconds, hold down 180, flushed after 240
  Output delay 8 milliseconds between packets
  Automatic network summarization is in effect
  Outgoing filter for all interfaces is
  Incoming filter for all interfaces is
  Default redistribution metric is 1
  Default version control: receive version 2, send version 2
      Interface Recv Send
      GigabitEthernet 2/42 2 2
      GigabitEthernet 2/41 2 2
      GigabitEthernet 2/31 2 2
      GigabitEthernet 2/11 2 2
Routing for Networks:
      10.300.10.0
      10.200.10.0
      10.11.20.0
      10.11.10.0

Routing Information Sources:
Gateway     Distance   Last Update
10.11.20.1  120        00:00:12

Distance: (default is 120)
Core2#
```

RIP Configuration on Core3

The following example shows how to configure RIPv2 on a host named Core3.
Example of Configuring RIPv2 on Core3

```
Core3(conf-if-gi-3/21)#router rip
Core3(conf-router_rip)#version 2
Core3(conf-router_rip)#network 192.168.1.0
Core3(conf-router_rip)#network 192.168.2.0
Core3(conf-router_rip)#network 10.11.30.0
Core3(conf-router_rip)#network 10.11.20.0
Core3(conf-router_rip)#show config
!
router rip
  network 10.0.0.0
  network 192.168.1.0
  network 192.168.2.0
  version 2
Core3(conf-router_rip)#
```

Core 3 RIP Output

The examples in this section show the core 2 RIP output.

- To display Core 3 RIP database, use the `show ip rip database` command.
- To display Core 3 RIP setup, use the `show ip route` command.
- To display Core 3 RIP activity, use the `show ip protocols` command.

Examples of the `show ip` Commands to View Learned RIP Routes on Core 3

The following example shows the `show ip rip database` command to view the learned RIP routes on Core 3.

```
Core3#show ip rip database
Total number of routes in RIP database: 7
10.11.10.0/24
    [120/1] via 10.11.20.2, 00:00:13, GigabitEthernet 3/21
10.200.10.0/24
    [120/1] via 10.11.20.2, 00:00:13, GigabitEthernet 3/21
10.300.10.0/24
    [120/1] via 10.11.20.2, 00:00:13, GigabitEthernet 3/21
10.11.20.0/24      directly connected,GigabitEthernet 3/21
10.11.30.0/24      directly connected,GigabitEthernet 3/11
10.0.0.0/8         auto-summary
192.168.1.0/24     directly connected,GigabitEthernet 3/43
192.168.1.0/24     auto-summary
192.168.2.0/24     directly connected,GigabitEthernet 3/44
192.168.2.0/24     auto-summary
Core3#
```

The following command shows the `show ip routes` command to view the RIP setup on Core 3.

```
Core3#show ip routes

Codes: C - connected, S - static, R - RIP,
       B - BGP, IN - internal BGP, EX - external BGP,LO - Locally Originated,
       O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
       N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
       E2 - OSPF external type 2, i - IS-IS, L1 - IS-IS level-1,
       L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
       > - non-active route, + - summary route

Gateway of last resort is not set
   Destination    Gateway           Dist/Metric Last Change
  ----------- ------- -----------          -----------
R  10.11.10.0/24  via 10.11.20.2, Gi 3/21   120/1      00:01:14
C  10.11.20.0/24  Direct, Gi 3/21           0/0        00:01:53
C  10.11.30.0/24  Direct, Gi 3/11           0/0        00:06:00
R  10.200.10.0/24 via 10.11.20.2, Gi        3/21 120/1 00:01:14
R  10.300.10.0/24 via 10.11.20.2, Gi        3/21 120/1 00:01:14
C  192.168.1.0/24 Direct, Gi                3/43 0/0   00:06:53
C  192.168.2.0/24 Direct, Gi                3/44 0/0   00:06:26
Core3#
```

The following example shows the `show ip protocols` command to show the RIP configuration activity on Core 3.

```
Core3#show ip protocols

Routing Protocol is "RIP"
  Sending updates every 30 seconds, next due in 6
  Invalid after 180 seconds, hold down 180, flushed after 240
  Output delay 8 milliseconds between packets
  Automatic network summarization is in effect
  Outgoing filter for all interfaces is
  Incoming filter for all interfaces is
  Default redistribution metric is 1
  Default version control: receive version 2, send version 2
      Interface Recv Send
      GigabitEthernet 3/21 2 2
      GigabitEthernet 3/11 2 2
      GigabitEthernet 3/44 2 2
      GigabitEthernet 3/43 2 2
Routing for Networks:
      10.11.20.0
      10.11.30.0
      192.168.2.0
      192.168.1.0

Routing Information Sources:
  Gateway     Distance  Last Update
  10.11.20.2  120       00:00:22

Distance: (default is 120)

Core3#
```

RIP Configuration Summary

Examples of Viewing RIP Configuration on Core 2 and Core 3

The following example shows viewing the RIP configuration on Core 2.

```
!
interface GigabitEthernet 2/11
  ip address 10.11.10.1/24
  no shutdown
!
interface GigabitEthernet 2/31
  ip address 10.11.20.2/24
  no shutdown
!
interface GigabitEthernet 2/41
  ip address 10.200.10.1/24
  no shutdown
!
interface GigabitEthernet 2/42
  ip address 10.250.10.1/24
  no shutdown
router rip
version 2
10.200.10.0
10.300.10.0
10.11.10.0
10.11.20.0
```

The following example shows viewing the RIP configuration on Core 3.

```
!
interface GigabitEthernet 3/11
  ip address 10.11.30.1/24
  no shutdown

!
interface GigabitEthernet 3/21
  ip address 10.11.20.1/24
  no shutdown

!
interface GigabitEthernet 3/43
  ip address 192.168.1.1/24
  no shutdown

!
interface GigabitEthernet 3/44
```

```
  ip address 192.168.2.1/24
  no shutdown

!
router rip
version 2
network 10.11.20.0
network 10.11.30.0
network 192.168.1.0
network 192.168.2.0
```

39

Remote Monitoring (RMON)

Remote monitoring (RMON) is supported on the S6000 platform.

RMON is an industry-standard implementation that monitors network traffic by sharing network monitoring information. RMON provides both 32-bit and 64-bit monitoring facility and long-term statistics collection on Dell Networking Ethernet interfaces.

RMON operates with the simple network management protocol (SNMP) and monitors all nodes on a local area network (LAN) segment. RMON monitors traffic passing through the router and segment traffic not destined for the router. The monitored interfaces may be chosen by using alarms and events with standard management information bases (MIBs).

Implementation Information

Configure SNMP prior to setting up RMON.

For a complete SNMP implementation description, refer to Simple Network Management Protocol (SNMP).

Configuring RMON requires using the RMON CLI and includes the following tasks:

- Setting the rmon Alarm
- Configuring an RMON Event
- Configuring RMON Collection Statistics
- Configuring the RMON Collection History

RMON implements the following standard request for comments (RFCs) (for more information, refer to the Standards Compliance chapter).

- RFC-2819
- RFC-3273
- RFC-3434
- RFC-4502

Fault Recovery

RMON provides the following fault recovery functions.

- **Interface Down** — When an RMON-enabled interface goes down, monitoring continues. However, all data values are registered as 0xFFFFFFFF (32 bits) or ixFFFFFFFFFFFFFFFF (64 bits). When the interface comes back up, RMON monitoring processes resumes.

  NOTE: A network management system (NMS) should be ready to interpret a down interface and plot the interface performance graph accordingly.
- **Line Card Down — The** same as Interface Down (see previous).
- **RPM Down**, **RPM Failover** — Master and standby route processor modules (RPMs) run the RMON sampling process in the background. Therefore, when an RPM goes down, the other RPM maintains

the sampled data — the new master RPM provides the same sampled data as did the old master — as long as the master RPM had been running long enough to sample all the data. NMS backs up all the long-term data collection and displays the failover downtime from the performance graph.

- **Chassis Down** — When a chassis goes down, all sampled data is lost. But the RMON configurations are saved in the configuration file. The sampling process continues after the chassis returns to operation.
- **Platform Adaptation** — RMON supports all Dell Networking chassis and all Dell Networking Ethernet interfaces.

Setting the rmon Alarm

To set an alarm on any MIB object, use the `rmon alarm` or `rmon hc-alarm` command in GLOBAL CONFIGURATION mode.

- Set an alarm on any MIB object.

  CONFIGURATION mode

  ```
  [no] rmon alarm number variable interval {delta | absolute} rising-threshold
  [value event-number] falling-threshold value event-number [owner string]
  ```

  OR

  ```
  [no] rmon hc-alarm number variable interval {delta | absolute} rising-
  threshold value event-number falling-threshold value event-number [owner
  string]
  ```

  Configure the alarm using the following optional parameters:

  - *`number`*: alarm number, an integer from 1 to 65,535, the value must be unique in the RMON Alarm Table.
  - *`variable`*: the MIB object to monitor — the variable must be in SNMP OID format; for example, 1.3.6.1.2.1.1.3. The object type must be a 32-bit integer for the `rmon alarm` command and 64 bits for the `rmon hc-alarm` command.
  - *`interval`*: time in seconds the alarm monitors the MIB variable, the value must be between 1 to 3,600.
  - `delta`: tests the change between MIB variables, this option is the alarmSampleType in the RMON Alarm table.
  - `absolute`: tests each MIB variable directly, this option is the alarmSampleType in the RMON Alarm table.
  - `rising-threshold` *`value`*: value at which the rising-threshold alarm is triggered or reset. For the `rmon alarm` command, this setting is a 32-bits value, for the `rmon hc-alarm` command, this setting is a 64-bits value.
  - *`event-number`*: event number to trigger when the rising threshold exceeds its limit. This value is identical to the alarmRisingEventIndex in the alarmTable of the RMON MIB. If there is no corresponding rising-threshold event, the value should be zero.
  - `falling-threshold` *`value`*: value at which the falling-threshold alarm is triggered or reset. For the `rmon alarm` command, this setting is a 32-bits value, for the `rmon hc-alarm` command this setting is a 64 bits value.
  - *`event-number`*: event number to trigger when the falling threshold exceeds its limit. This value is identical to the alarmFallingEventIndex in the alarmTable of the RMON MIB. If there is no corresponding falling-threshold event, the value should be zero.
  - `owner` *`string`*: (Optional) specifies an owner for the alarm, this setting is the alarmOwner object in the alarmTable of the RMON MIB. Default is a **null-terminated string**.

Example of the `rmon alarm` Command

To disable the alarm, use the `no` form of the command.

The following example configures RMON alarm number 10. The alarm monitors the MIB variable 1.3.6.1.2.1.2.2.1.20.1 (ifEntry.ifOutErrors) once every 20 seconds until the alarm is disabled, and checks the rise or fall of the variable. The alarm is triggered when the 1.3.6.1.2.1.2.2.1.20.1 value shows a MIB counter increase of 15 or more (such as from 100000 to 100015). The alarm then triggers event number 1, which is configured with the RMON event command. Possible events include a log entry or an SNMP trap. If the 1.3.6.1.2.1.2.2.1.20.1 value changes to 0 (falling-threshold 0), the alarm is reset and can be triggered again.

```
Dell(conf)#rmon alarm 10 1.3.6.1.2.1.2.2.1.20.1 20 delta rising-threshold
15 1 falling-threshold 0 1 owner nms1
```

Configuring an RMON Event

To add an event in the RMON event table, use the `rmon event` command in GLOBAL CONFIGURATION mode.

- Add an event in the RMON event table.

  CONFIGURATION mode

  ```
  [no] rmon event number [log] [trap community] [description string] [owner
  string]
  ```

  - *`number`*: assign an event number in integer format from 1 to 65535. The number value must be unique in the RMON event table.
  - *`log`*: (Optional) enter the keyword log to generate an RMON event log, it sets the eventType to either log or log-and-snmptrap in the RMON event table. The default is None.
  - `trap` *`community`*: (Optional) enter the keyword trap and SNMP community string to generate SNMP traps for an RMON event entry, it sets the eventType to either snmptrap or log-and-snmptrap in the RMON event table. This option generates syslog as well along with SNMP traps.
  - `description` *`string`*: (Optional) specifies a description of the event, which is identical to the event description in the eventTable of the RMON MIB. The default is a **null-terminated string**.
  - `owner` *`string`*: (Optional) owner of this event, which is identical to the eventOwner in the eventTable of the RMON MIB. Default is a **null-terminated string**.

Example of the `rmon event` Command

To disable RMON on the interface, use the `no` form of this command.

In the following example, the configuration creates RMON event number 1, with the description "High ifOutErrors", and generates a log entry when an alarm triggers the event. The user *nms1* owns the row that is created in the event table by this command. This configuration also generates an SNMP trap when the event is triggered using the SNMP community string "eventtrap".

```
Dell(conf)#rmon event 1 log trap eventtrap description "High ifOutErrors" owner
nms1
```

Configuring RMON Collection Statistics

To enable RMON MIB statistics collection on an interface, use the `RMON collection statistics` command in INTERFACE CONFIGURATION mode.

- Enable RMON MIB statistics collection.

  CONFIGURATION INTERFACE (config-if) mode

```
[no] rmon collection statistics {controlEntry integer} [owner ownername]
```

  - `controlEntry`: specifies the RMON group of statistics using a value.
  - *`integer`*: a value from 1 to 65,535 that identifies the RMON Statistics Table. The value must be unique in the RMON Statistic Table.
  - `owner`: (Optional) specifies the name of the owner of the RMON group of statistics.
  - *`ownername`*: (Optional) records the name of the owner of the RMON group of statistics. The default is a **null-terminated string**.

Example of the `rmon collection statistics` Command

To remove a specified RMON statistics collection, use the `no` form of this command.

The following command example enables the RMON statistics collection on the interface, with an ID value of 20 and an owner of *john*.

```
Dell(conf-if-mgmt)#rmon collection statistics controlEntry 20 owner john
```

Configuring the RMON Collection History

To enable the RMON MIB history group of statistics collection on an interface, use the `rmon collection history` command in INTERFACE CONFIGURATION mode.

- Configure the RMON MIB history group of statistics collection.

  CONFIGURATION INTERFACE (config-if) mode

```
[no] rmon collection history {controlEntry integer} [owner ownername]
[buckets bucket-number] [interval seconds]
```

  - `controlEntry`: specifies the RMON group of statistics using a value.
  - *`integer`*: a value from 1 to 65,535 that identifies the RMON group of statistics. The value must be a unique index in the RMON History Table.
  - `owner`: (Optional) specifies the name of the owner of the RMON group of statistics. The default is a **null-terminated string**.
  - *`ownername`*: (Optional) records the name of the owner of the RMON group of statistics.
  - `buckets`: (Optional) specifies the maximum number of buckets desired for the RMON collection history group of statistics.
  - *`bucket-number`*: (Optional) a value associated with the number of buckets specified for the RMON collection history group of statistics. The value is limited to from 1 to 1000. The default is **50** (as defined in RFC-2819).
  - `interval`: (Optional) specifies the number of seconds in each polling cycle.
  - `seconds`: (Optional) the number of seconds in each polling cycle. The value is ranged from 5 to 3,600 (Seconds). The default is **1,800** (as defined in RFC-2819).

Example of the `rmon collection history` Command

To remove a specified RMON history group of statistics collection, use the `no` form of this command.

The following command example enables an RMON MIB collection history group of statistics with an ID number of 20 and an owner of *john*, both the sampling interval and the number of buckets use their respective defaults.

```
Dell(conf-if-mgmt)#rmon collection history controlEntry 20 owner john
```

40

Rapid Spanning Tree Protocol (RSTP)

Rapid spanning tree protocol (RSTP) is supported on the S6000 platform.

Protocol Overview

RSTP is a Layer 2 protocol — specified by IEEE 802.1w — that is essentially the same as spanning-tree protocol (STP) but provides faster convergence and interoperability with switches configured with STP and multiple spanning tree protocol (MSTP).

The Dell Networking OS supports three other variations of spanning tree, as shown in the following table.

Table 62. Spanning Tree Variations Dell Networking OS Supports

Dell Networking Term	IEEE Specification
Spanning Tree Protocol (STP)	802.1d
Rapid Spanning Tree Protocol (RSTP)	802.1w
Multiple Spanning Tree Protocol (MSTP)	802.1s
Per-VLAN Spanning Tree Plus (PVST+)	Third Party

Configuring Rapid Spanning Tree

Configuring RSTP is a two-step process.

1. Configure interfaces for Layer 2.
2. Enable the rapid spanning tree protocol.

Related Configuration Tasks

- Adding and Removing Interfaces
- Modifying Global Parameters
- Modifying Interface Parameters
- Configuring an EdgePort
- Prevent Network Disruptions with BPDU Guard
- Influencing RSTP Root Selection
- Configuring Spanning Trees as Hitless
- Enabling SNMP Traps for Root Elections and Topology Changes
- Configuring Fast Hellos for Link State Detection
- Flush MAC Addresses after a Topology Change

Important Points to Remember

- RSTP is disabled by default.
- Dell Networking OS supports only one Rapid Spanning Tree (RST) instance.
- All interfaces in virtual local area networks (VLANs) and all enabled interfaces in Layer 2 mode are automatically added to the RST topology.
- Adding a group of ports to a range of VLANs sends multiple messages to the rapid spanning tree protocol (RSTP) task, avoid using the `range` command. When using the `range` command, Dell Networking recommends limiting the range to five ports and 40 VLANs.

RSTP and VLT

Virtual link trunking (VLT) provides loop-free redundant topologies and does not require RSTP.

RSTP can cause temporary port state blocking and may cause topology changes after link or node failures. Spanning tree topology changes are distributed to the entire Layer 2 network, which can cause a network-wide flush of learned media access control (MAC) and address resolution protocol (ARP) addresses, requiring these addresses to be re-learned. However, enabling RSTP can detect potential loops caused by non-system issues such as cabling errors or incorrect configurations. RSTP is useful for potential loop detection but to minimize possible topology changes after link or node failure, configure it using the following specifications.

The following recommendations help you avoid these issues and the associated traffic loss caused by using RSTP when you enable VLT on both VLT peers:

- Configure any ports at the edge of the spanning tree's operating domain as edge ports, which are directly connected to end stations or server racks. Ports connected directly to Layer 3-only routers not running STP should have RSTP disabled or be configured as edge ports.
- Ensure that the primary VLT node is the root bridge and the secondary VLT peer node has the second-best bridge ID in the network. If the primary VLT peer node fails, the secondary VLT peer node becomes the root bridge, avoiding problems with spanning tree port state changes that occur when a VLT node fails or recovers.
- Even with this configuration, if the node has non-VLT ports using RSTP that are not configured as edge ports and are connected to other layer 2 switches, spanning tree topology changes can still be detected after VLT node recovery. To avoid this scenario, ensure that you configure any non-VLT ports as edge ports or have RSTP disabled.

Configuring Interfaces for Layer 2 Mode

To configure and enable interfaces in Layer 2 mode, use the following commands.
All interfaces on all bridges that participate in Rapid Spanning Tree must be in Layer 2 and enabled.

1. If the interface has been assigned an IP address, remove it.
   INTERFACE mode

   ```
   no ip address
   ```
2. Place the interface in Layer 2 mode.
   INTERFACE mode

   ```
   switchport
   ```
3. Enable the interface.

INTERFACE mode

```
no shutdown
```

Example of Verifying an Interface is in Layer 2 Mode and Enabled

To verify that an interface is in Layer 2 mode and enabled, use the `show config` command from INTERFACE mode. The bold lines indicate that the interface is in Layer 2 mode.

```
Dell(conf-if-gi-1/1)#show config
!
interface GigabitEthernet 1/1
  no ip address
  switchport
no shutdown
Dell(conf-if-gi-1/1)#
```

Enabling Rapid Spanning Tree Protocol Globally

Enable RSTP globally on all participating bridges; it is not enabled by default.

When you enable RSTP, all physical and port-channel interfaces that are enabled and in Layer 2 mode are automatically part of the RST topology.

- Only one path from any bridge to any other bridge is enabled.
- Bridges block a redundant path by disabling one of the link ports.

To enable RSTP globally for all Layer 2 interfaces, use the following commands.

1. Enter PROTOCOL SPANNING TREE RSTP mode.
   CONFIGURATION mode

   ```
   protocol spanning-tree rstp
   ```
2. Enable RSTP.
   PROTOCOL SPANNING TREE RSTP mode

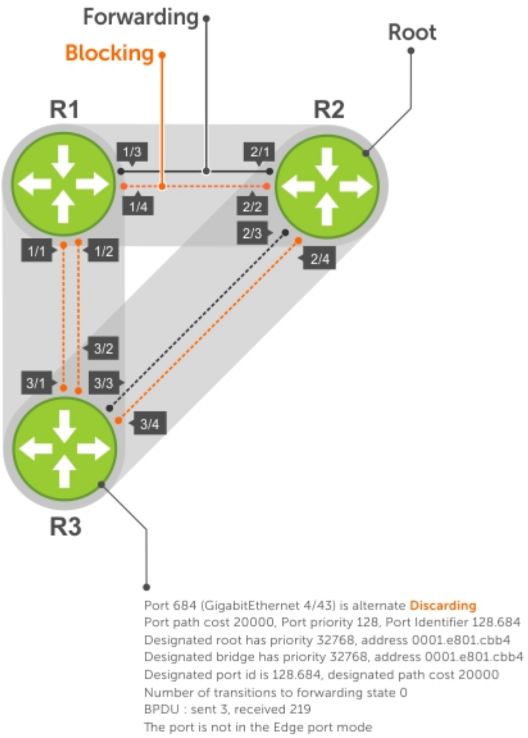
   ```
   no disable
   ```

Examples of the RSTP `show` Commands

To disable RSTP globally for all Layer 2 interfaces, enter the `disable` command from PROTOCOL SPANNING TREE RSTP mode.

To verify that RSTP is enabled, use the `show config` command from PROTOCOL SPANNING TREE RSTP mode. The bold line indicates that RSTP is enabled.

```
Dell(conf-rstp)#show config
!
protocol spanning-tree rstp
no disable
Dell(conf-rstp)#
```



```
Port 684 (GigabitEthernet 4/43) is alternate Discarding
Port path cost 20000, Port priority 128, Port Identifier 128.684
Designated root has priority 32768, address 0001.e801.cbb4
Designated bridge has priority 32768, address 0001.e801.cbb4
Designated port id is 128.684, designated path cost 20000
Number of transitions to forwarding state 0
BPDU : sent 3, received 219
The port is not in the Edge port mode
```

Figure 103. Rapid Spanning Tree Enabled Globally

To view the interfaces participating in RSTP, use the `show spanning-tree rstp` command from EXEC privilege mode. If a physical interface is part of a port channel, only the port channel is listed in the command output.

```
Dell#show spanning-tree rstp
Root Identifier has priority 32768, Address 0001.e801.cbb4
Root Bridge hello time 2, max age 20, forward delay 15, max hops 0
Bridge Identifier has priority 32768, Address 0001.e801.cbb4
Configured hello time 2, max age 20, forward delay 15, max hops 0
We are the root
Current root has priority 32768, Address 0001.e801.cbb4
Number of topology changes 4, last change occurred 00:02:17 ago on Gi 1/26

Port 377 (GigabitEthernet 2/1) is designated Forwarding
Port path cost 20000, Port priority 128, Port Identifier 128.377
Designated root has priority 32768, address 0001.e801.cbb4
Designated bridge has priority 32768, address 0001.e801.cbb4
Designated port id is 128.377, designated path cost 0
Number of transitions to forwarding state 1
BPDU : sent 121, received 9
The port is not in the Edge port mode

Port 378 (GigabitEthernet 2/2) is designated Forwarding
Port path cost 20000, Port priority 128, Port Identifier 128.378
Designated root has priority 32768, address 0001.e801.cbb4
Designated bridge has priority 32768, address 0001.e801.cbb4
Designated port id is 128.378, designated path cost 0
Number of transitions to forwarding state 1
```

```
BPDU : sent 121, received 2
The port is not in the Edge port mode

Port 379 (GigabitEthernet 2/3) is designated Forwarding
Port path cost 20000, Port priority 128, Port Identifier 128.379
Designated root has priority 32768, address 0001.e801.cbb4
Designated bridge has priority 32768, address 0001.e801.cbb4
Designated port id is 128.379, designated path cost 0
Number of transitions to forwarding state 1
BPDU : sent 121, received 5
The port is not in the Edge port mode

Port 380 (GigabitEthernet 2/4) is designated Forwarding
Port path cost 20000, Port priority 128, Port Identifier 128.380
Designated root has priority 32768, address 0001.e801.cbb4
Designated bridge has priority 32768, address 0001.e801.cbb4
Designated port id is 128.380, designated path cost 0

Number of transitions to forwarding state 1
BPDU : sent 147, received 3
The port is not in the Edge port mode
```

To confirm that a port is participating in RSTP, use the `show spanning-tree rstp brief` command from EXEC privilege mode.

```
R3#show spanning-tree rstp brief
Executing IEEE compatible Spanning Tree Protocol
Root ID Priority 32768, Address 0001.e801.cbb4
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID Priority 32768, Address 0001.e80f.1dad
Configured hello time 2, max age 20, forward delay 15
Interface                             Designated
Name    PortID  Prio Cost   Sts Cost  Bridge ID             PortID
---------- -------- ----   ------- --- ------- ------------- --------
Gi 3/1  128.681 128  20000  BLK 20000 32768 0001.e80b.88bd  128.469
Gi 3/2  128.682 128  20000  BLK 20000 32768 0001.e80b.88bd  128.470
Gi 3/3  128.683 128  20000  FWD 20000 32768 0001.e801.cbb4  128.379
Gi 3/4  128.684 128  20000  BLK 20000 32768 0001.e801.cbb4  128.380
Interface
Name    Role  PortID  Prio Cost  Sts Cost   Link-type Edge
------- --- ------ -------- ---- ------- --- -----------
Gi 3/1  Altr  128.681 128 20000  BLK 20000  P2P       No
Gi 3/2  Altr  128.682 128 20000  BLK 20000  P2P       No
Gi 3/3  Root  128.683 128 20000  FWD 20000  P2P       No
Gi 3/4  Altr  128.684 128 20000  BLK 20000  P2P       No
R3#
```

Adding and Removing Interfaces

To add and remove interfaces, use the following commands.
To add an interface to the Rapid Spanning Tree topology, configure it for Layer 2 and it is automatically added. If you previously disabled RSTP on the interface using the command `no spanning-tree 0` command, re-enable it using the `spanning-tree 0` command.

- Remove an interface from the Rapid Spanning Tree topology.

  `no spanning-tree 0`

Modifying Global Parameters

You can modify RSTP parameters.
The root bridge sets the values for forward-delay, hello-time, and max-age and overwrites the values set on other bridges participating in the Rapid Spanning Tree group.

- **Forward-delay** — the amount of time an interface waits in the Listening state and the Learning state before it transitions to the Forwarding state.
- **Hello-time** — the time interval in which the bridge sends RSTP BPDUs.
- **Max-age** — the length of time the bridge maintains configuration information before it refreshes that information by recomputing the RST topology.

NOTE: Dell Networking recommends that only experienced network administrators change the Rapid Spanning Tree group parameters. Poorly planned modification of the RSTP parameters can negatively affect network performance.

The following table displays the default values for RSTP.

Table 63. RSTP Default Values

RSTP Parameter	Default Value
Forward Delay	**15 seconds**
Hello Time	**2 seconds**
Max Age	**20 seconds**
Port Cost:	Port Cost:
• 100-Mb/s Ethernet interfaces	• **200000**
• 1-Gigabit Ethernet interfaces	• **20000**
• 10-Gigabit Ethernet interfaces	• **2000**
• Port Channel with 100 Mb/s Ethernet interfaces	• **180000**
• Port Channel with 1-Gigabit Ethernet interfaces	• **18000**
• Port Channel with 10-Gigabit Ethernet interfaces	• **1800**
Port Priority	**128**

To change these parameters, use the following commands.

- Change the forward-delay parameter.

  PROTOCOL SPANNING TREE RSTP mode

  ```
  forward-delay seconds
  ```

  The range is from 4 to 30.

  The default is **15 seconds**.
- Change the hello-time parameter.

  PROTOCOL SPANNING TREE RSTP mode

  ```
  hello-time seconds
  ```

NOTE: With large configurations (especially those configurations with more ports) Dell Networking recommends increasing the hello-time.

The range is from 1 to 10.

The default is **2 seconds**.

- Change the max-age parameter.

  PROTOCOL SPANNING TREE RSTP mode

  ```
  max-age seconds
  ```

  The range is from 6 to 40.

  The default is **20 seconds**.

To view the current values for global parameters, use the `show spanning-tree rstp` command from EXEC privilege mode.

Enabling SNMP Traps for Root Elections and Topology Changes

To enable SNMP traps, use the following command.

- Enable SNMP traps for RSTP, MSTP, and PVST+ collectively.

  ```
  snmp-server enable traps xstp
  ```

Modifying Interface Parameters

On interfaces in Layer 2 mode, you can set the port cost and port priority values.

- **Port cost** — a value that is based on the interface type. The previous table lists the default values. The greater the port cost, the less likely the port is selected to be a forwarding port.
- **Port priority** — influences the likelihood that a port is selected to be a forwarding port in case that several ports have the same port cost.

To change the port cost or priority of an interface, use the following commands.

- Change the port cost of an interface.

  INTERFACE mode

  ```
  spanning-tree rstp cost cost
  ```

  The range is from 0 to 65535.

  The default is listed in the previous table.
- Change the port priority of an interface.

  INTERFACE mode

  ```
  spanning-tree rstp priority priority-value
  ```

  The range is from 0 to 15.

  The default is **128**.

To view the current values for interface parameters, use the `show spanning-tree rstp` command from EXEC privilege mode.

Enabling SNMP Traps for Root Elections and Topology Changes

To enable SNMP traps collectively, use this command.

Enable SNMP traps for RSTP, MSTP, and PVST+ collectively.

```
snmp-server enable traps xstp
```

Influencing RSTP Root Selection

RSTP determines the root bridge, but you can assign one bridge a lower priority to increase the likelihood that it is selected as the root bridge.
To change the bridge priority, use the following command.

- Assign a number as the bridge priority or designate it as the primary or secondary root.
  PROTOCOL SPANNING TREE RSTP mode

  ```
  bridge-priority priority-value
  ```

  - *`priority-value`* The range is from 0 to 65535. The lower the number assigned, the more likely this bridge becomes the root bridge.

  The default is **32768**. Entries must be multiples of 4096.

Example of the bridge-priority Command

A console message appears when a new root bridge has been assigned. The following example example shows the console message after the `bridge-priority` command is used to make R2 the root bridge (shown in bold).

```
Dell(conf-rstp)#bridge-priority 4096
04:27:59: %RPM0-P:RP2 %SPANMGR-5-STP_ROOT_CHANGE: RSTP root changed. My Bridge
ID:
4096:0001.e80b.88bd Old Root: 32768:0001.e801.cbb4 New Root: 4096:0001.e80b.88bd
```

Configuring an EdgePort

The EdgePort feature enables interfaces to begin forwarding traffic approximately 30 seconds sooner. In this mode an interface forwards frames by default until it receives a BPDU that indicates that it should behave otherwise; it does not go through the Learning and Listening states. The `bpduguard shutdown-on-violation` option causes the interface hardware to be shut down when it receives a BPDU. When only `bpduguard` is implemented, although the interface is placed in an Error Disabled state when receiving the BPDU, the physical interface remains up and spanning-tree drops packets in the hardware after a BPDU violation. BPDUs are dropped in the software after receiving the BPDU violation. This feature is the same as PortFast mode in Spanning Tree.

⚠ **CAUTION: Configure EdgePort only on links connecting to an end station. If you enable EdgePort on an interface connected to a network, it can cause loops.**

Dell Networking OS Behavior: Regarding **bpduguard shutdown-on-violation** behavior:

- If the interface to be shut down is a port channel, all the member ports are disabled in the hardware.
- When you add a physical port to a port channel already in the Error Disable state, the new member port is also disabled in the hardware.
- When you remove a physical port from a port channel in the Error Disable state, the error disabled state is cleared on this physical port (the physical port is enabled in the hardware).
- The `reset linecard` command does not clear the Error Disabled state of the port or the hardware disabled state. The interface continues to be disables in the hardware.
- You can clear the Error Disabled state with any of the following methods:
  - Perform an `shutdown` command on the interface.
  - Disable the `shutdown-on-violation` command on the interface (the `no spanning-tree stp-id portfast [bpduguard | [shutdown-on-violation]]` command).
  - Disable spanning tree on the interface (the `no spanning-tree` command in INTERFACE mode).
  - Disable global spanning tree (the `no spanning-tree` command in CONFIGURATION mode).

To enable EdgePort on an interface, use the following command.

- Enable EdgePort on an interface.

  INTERFACE mode

  ```
  spanning-tree rstp edge-port [bpduguard | shutdown-on-violation]
  ```

Example of Verifying an EdgePort is Enabled on an Interface

To verify that EdgePort is enabled on a port, use the `show spanning-tree rstp` command from EXEC privilege mode or the `show config` command from INTERFACE mode.

NOTE: Dell Networking recommends using the `show config` command from INTERFACE mode.

In the following example, the bold line indicates that the interface is in EdgePort mode.

```
Dell(conf-if-gi-2/0)#show config
!
interface GigabitEthernet 2/0
  no ip address
  switchport
  spanning-tree rstp edge-port
  shutdown
Dell(conf-if-gi-2/0)#
```

Configuring Fast Hellos for Link State Detection

Fast hellos for link state detection are available on the S6000 platform.
To achieve sub-second link-down detection so that convergence is triggered faster, use RSTP fast hellos. The standard RSTP link-state detection mechanism does not offer the same low link-state detection speed.

RSTP fast hellos decrease the hello interval to the order of milliseconds and all timers derived from the hello timer are adjusted accordingly. This feature does not inter-operate with other vendors, and is available only for RSTP.

- Configure a hello time on the order of milliseconds.

  PROTOCOL RSTP mode

  ```
  hello-time milli-second interval
  ```

The range is from 50 to 950 milliseconds.

Example of Verifying Hello-Time Interval

```
Dell(conf-rstp)#do show spanning-tree rstp brief
Executing IEEE compatible Spanning Tree Protocol
  Root ID     Priority 0, Address 0001.e811.2233
  Root Bridge hello time 50 ms, max age 20, forward delay 15
  Bridge ID   Priority 0, Address 0001.e811.2233
  We are the root
  Configured hello time 50 ms, max age 20, forward delay 15
```

NOTE: The hello time is encoded in BPDUs in increments of 1/256ths of a second. The standard minimum hello time in seconds is 1 second, which is encoded as 256. Millisecond. hello times are encoded using values less than 256; the millisecond hello time equals (x/1000)*256. When you configure millisecond hellos, the default hello interval of 2 seconds is still used for edge ports; the millisecond hello interval is not used.

41

Software-Defined Networking (SDN)

Dell Networking operating software supports Software-Defined Networking (SDN). For more information, refer to the *SDN Deployment Guide*.

42

Service Provider Bridging

Service provider bridging is supported on the S6000 platform.

VLAN Stacking

Virtual local area network (VLAN) stacking is supported on the S6000 platform.

VLAN stacking, also called Q-in-Q, is defined in IEEE 802.1ad — Provider *Bridges*, which is an amendment to IEEE 802.1Q — Virtual *Bridged Local Area Networks*. It enables service providers to use 802.1Q architecture to offer separate VLANs to customers with no coordination between customers, and minimal coordination between customers and the provider.

Using only 802.1Q VLAN tagging all customers would have to use unique VLAN IDs to ensure that traffic is segregated, and customers and the service provider would have to coordinate to ensure that traffic mapped correctly across the provider network. Even under ideal conditions, customers and the provider would still share the 4094 available VLANs.

Instead, 802.1ad allows service providers to add their own VLAN tag to frames traversing the provider network. The provider can then differentiate customers even if they use the same VLAN ID, and providers can map multiple customers to a single VLAN to overcome the 4094 VLAN limitation. Forwarding decisions in the provider network are based on the provider VLAN tag only, so the provider can map traffic through the core independently; the customer and provider only coordinate at the provider edge.

At the access point of a VLAN-stacking network, service providers add a VLAN tag, the S-Tag, to each frame before the 802.1Q tag. From this point, the frame is double-tagged. The service provider uses the S-Tag, to forward the frame traffic across its network. At the egress edge, the provider removes the S-Tag, so that the customer receives the frame in its original condition, as shown in the following illustration.

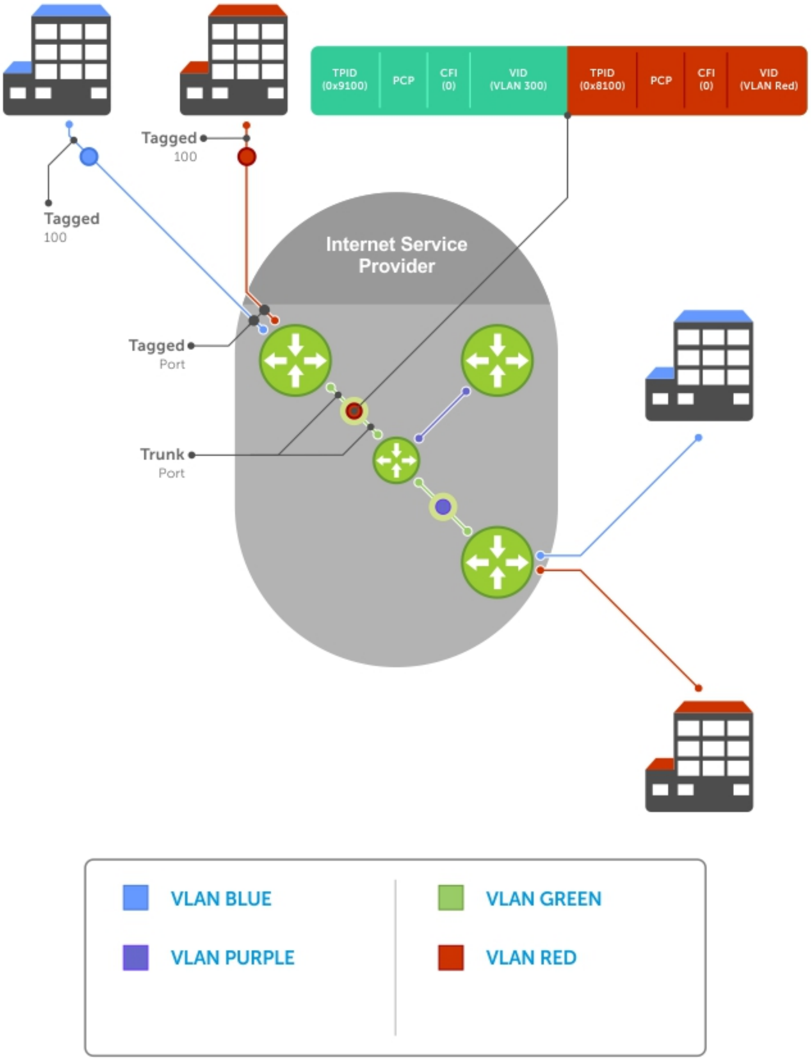

TPID (0x9100)
PCP
CFI (0)
VID (VLAN 300)
TPID (0x8100)
PCP
CFI (0)
VID (VLAN Red)
Tagged 100
Tagged 100
Internet Service Provider
Tagged Port
Trunk Port
VLAN BLUE
VLAN GREEN
VLAN PURPLE
VLAN RED

Figure 104. VLAN Stacking in a Service Provider Network

Important Points to Remember

- Interfaces that are members of the Default VLAN and are configured as VLAN-Stack access or trunk ports do not switch untagged traffic. To switch traffic, add these interfaces to a non-default VLAN-Stack-enabled VLAN.
- Dell Networking cautions against using the same MAC address on different customer VLANs, on the same VLAN-Stack VLAN.
- This limitation becomes relevant if you enable the port as a multi-purpose port (carrying single-tagged and double-tagged traffic).

Configure VLAN Stacking

Configuring VLAN-Stacking is a three-step process.

1. Creating Access and Trunk Ports
2. Assign access and trunk ports to a VLAN (Creating Access and Trunk Ports).
3. Enabling VLAN-Stacking for a VLAN.

Related Configuration Tasks

- Configuring the Protocol Type Value for the Outer VLAN Tag
- Configuring Dell Networking OS Options for Trunk Ports
- Debugging VLAN Stacking
- VLAN Stacking in Multi-Vendor Networks

Creating Access and Trunk Ports

To create access and trunk ports, use the following commands.

- **Access port** — a port on the service provider edge that directly connects to the customer. An access port may belong to only one service provider VLAN.
- **Trunk port** — a port on a service provider bridge that connects to another service provider bridge and is a member of multiple service provider VLANs.

Physical ports and port-channels can be access or trunk ports.

1. Assign the role of access port to a Layer 2 port on a provider bridge that is connected to a customer.

   INTERFACE mode

   ```
   vlan-stack access
   ```
2. Assign the role of trunk port to a Layer 2 port on a provider bridge that is connected to another provider bridge.

   INTERFACE mode

   ```
   vlan-stack trunk
   ```
3. Assign all access ports and trunk ports to service provider VLANs.

   INTERFACE VLAN mode

   ```
   member
   ```

Example of Displaying the VLAN-Stack Configuration for a Switchport

To display the VLAN-Stacking configuration for a switchport, use the `show config` command from INTERFACE mode.

```
Dell#show run interface gi 7/0
!
interface GigabitEthernet 7/0
  no ip address
  switchport
  vlan-stack access
  no shutdown
Dell#show run interface gi 7/12
!
```

```
interface GigabitEthernet 7/12
  no ip address
  switchport
  vlan-stack trunk
  no shutdown
```

Enable VLAN-Stacking for a VLAN

To enable VLAN-Stacking for a VLAN, use the following command.

- Enable VLAN-Stacking for the VLAN.

  INTERFACE VLAN mode

  ```
  vlan-stack compatible
  ```

Example of Viewing VLAN Stack Member Status

To display the status and members of a VLAN, use the `show vlan` command from EXEC Privilege mode. Members of a VLAN-Stacking-enabled VLAN are marked with an M in column Q.

```
Dell#show vlan

Codes: * - Default VLAN, G - GVRP VLANs

NUM  Status      Q Ports
* 1  Active      U Gi 13/0-5,18
  2  Inactive
  3  Inactive
  4  Inactive
  5  Inactive
  6  Active      M Po1(Gi 13/14-15)
                 M Gi 13/13
Dell#
```

Configuring the Protocol Type Value for the Outer VLAN Tag

The tag protocol identifier (TPID) field of the S-Tag is user-configurable.
To set the S-Tag TPID, use the following command.

- Select a value for the S-Tag TPID.

  CONFIGURATION mode

  ```
  vlan-stack protocol-type
  ```

  The default is **9100**.

To display the S-Tag TPID for a VLAN, use the `show running-config` command from EXEC privilege mode. Dell Networking OS displays the S-Tag TPID only if it is a non-default value.

Configuring Dell Networking OS Options for Trunk Ports

802.1ad trunk ports may also be tagged members of a VLAN so that it can carry single and double-tagged traffic.
You can enable trunk ports to carry untagged, single-tagged, and double-tagged VLAN traffic by making the trunk port a hybrid port.

To configure trunk ports, use the following commands.

1. Configure a trunk port to carry untagged, single-tagged, and double-tagged traffic by making it a hybrid port.
   INTERFACE mode

   ```
   portmode hybrid
   ```
2. Add the port to a 802.1Q VLAN as tagged or untagged.
   INTERFACE VLAN mode

   ```
   [tagged | untagged]
   ```

Example of Configuring a Trunk Port as a Hybrid Port and Adding it to Stacked VLANs

In the following example, GigabitEthernet 0/1 is a trunk port that is configured as a hybrid port and then added to VLAN 100 as untagged VLAN 101 as tagged, and VLAN 103, which is a stacking VLAN.

```
Dell(conf)#int gi 0/1
Dell(conf-if-gi-0/1)#portmode hybrid
Dell(conf-if-gi-0/1)#switchport
Dell(conf-if-gi-0/1)#vlan-stack trunk
Dell(conf-if-gi-0/1)#show config
!
interface GigabitEthernet 0/1
  no ip address
  portmode hybrid
  switchport
  vlan-stack trunk
  shutdown
Dell(conf-if-gi-0/1)#interface vlan 100
Dell(conf-if-vl-100)#untagged gigabitethernet 0/1
Dell(conf-if-vl-100)#interface vlan 101
Dell(conf-if-vl-101)#tagged gigabitethernet 0/1
Dell(conf-if-vl-101)#interface vlan 103
Dell(conf-if-vl-103)#vlan-stack compatible
Dell(conf-if-vl-103-stack)#member gigabitethernet 0/1
Dell(conf-if-vl-103-stack)#do show vlan

Codes: * - Default VLAN, G - GVRP VLANs
Q: U - Untagged, T - Tagged
   x - Dot1x untagged, X - Dot1x tagged
   G - GVRP tagged, M - Vlan-stack

NUM    Status     Description  Q Ports
* 1    Inactive
  100  Inactive                U Gi 0/1
  101  Inactive                T Gi 0/1
  103  Inactive                M Gi 0/1
```

Debugging VLAN Stacking

To debug VLAN stacking, use the following command.

- Debug the internal state and membership of a VLAN and its ports.
  `debug member`

Example of Debugging a VLAN and its Ports

The port notations are as follows:

- **MT** — stacked trunk
- **MU** — stacked access port
- **T** — 802.1Q trunk port
- **U** — 802.1Q access port
- **NU** — Native VLAN (untagged)

```
Dell# debug member vlan 603
vlan id  : 603
ports    : Gi 2/47 (MT), Gi 3/1(MU), Gi 3/25(MT), Gi 3/26(MT), Gi 3/27(MU)

Dell#debug member port gigabitethernet 2/47
vlan id  : 603 (MT), 100(T), 101(NU)
Dell#
```

VLAN Stacking in Multi-Vendor Networks

The first field in the VLAN tag is the tag protocol identifier (TPID), which is 2 bytes. In a VLAN-stacking network, after the frame is double tagged, the outer tag TPID must match the TPID of the next-hop system.

While 802.1Q requires that the inner tag TPID is 0x8100, it does not require a specific value for the outer tag TPID. Systems may use any 2-byte value; Dell Networking OS uses 0x9100 (shown in the following) while non-Dell Networking systems might use a different value.

If the next-hop system’s TPID does not match the outer-tag TPID of the incoming frame, the system drops the frame. For example, as shown in the following, the frame originating from Building A is tagged VLAN RED, and then double-tagged VLAN PURPLE on egress at R4. The TPID on the outer tag is 0x9100. R2’s TPID must also be 0x9100, and it is, so R2 forwards the frame.

Given the matching-TPID requirement, there are limitations when you employ Dell Networking systems at network edges, at which, frames are either double tagged on ingress (R4) or the outer tag is removed on egress (R3).

VLAN Stacking

The default TPID for the outer VLAN tag is 0x9100. The system allows you to configure both bytes of the 2 byte TPID.

Previous versions allowed you to configure the first byte only, and thus, the systems did not differentiate between TPIDs with a common first byte. For example, 0x8100 and any other TPID beginning with 0x81 were treated as the same TPID, as shown in the following illustration. Dell Networking OS Versions 8.2.1.0 and later differentiate between 0x9100 and 0x91XY, also shown in the following illustration.

You can configure the first 8 bits of the TPID using the `vlan-stack protocol-type` command.

The TPID is global. Ingress frames that do not match the system TPID are treated as untagged. This rule applies for both the outer tag TPID of a double-tagged frame and the TPID of a single-tagged frame.

For example, if you configure TPID 0x9100, the system treats 0x8100 and untagged traffic the same and maps both types to the default VLAN, as shown by the frame originating from Building C. For the same traffic types, if you configure TPID 0x8100, the system is able to differentiate between 0x8100 and untagged traffic and maps each to the appropriate VLAN, as shown by the packet originating from Building A.

Therefore, a mismatched TPID results in the port not differentiating between tagged and untagged traffic.

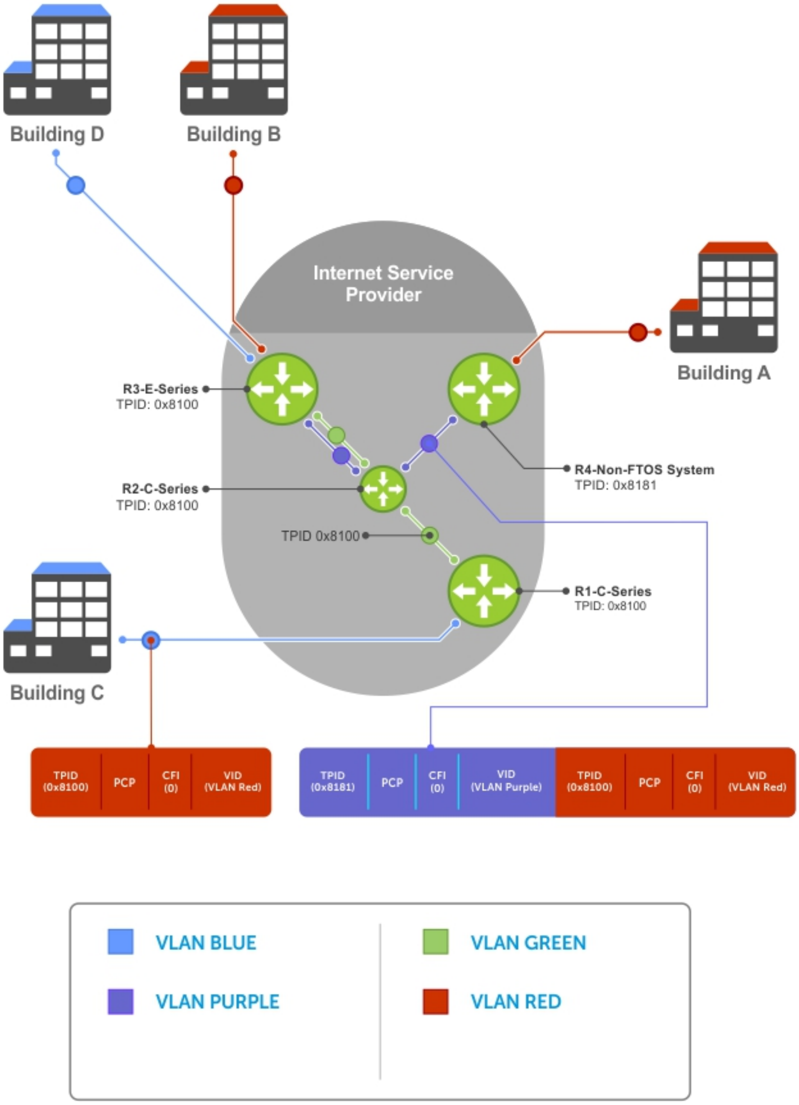

Building D
Building B
Internet Service Provider
Building A
R3-E-Series
TPID: 0x8100
R4-Non-FTOS System
TPID: 0x8181
R2-C-Series
TPID: 0x8100
TPID 0x8100
R1-C-Series
TPID: 0x8100
Building C
TPID (0x8100)
PCP
CFI (0)
VID (VLAN Red)
TPID (0x8181)
PCP
CFI (0)
VID (VLAN Purple)
TPID (0x8100)
PCP
CFI (0)
VID (VLAN Red)
VLAN BLUE
VLAN GREEN
VLAN PURPLE
VLAN RED

Figure 105. Single and Double-Tag TPID Match

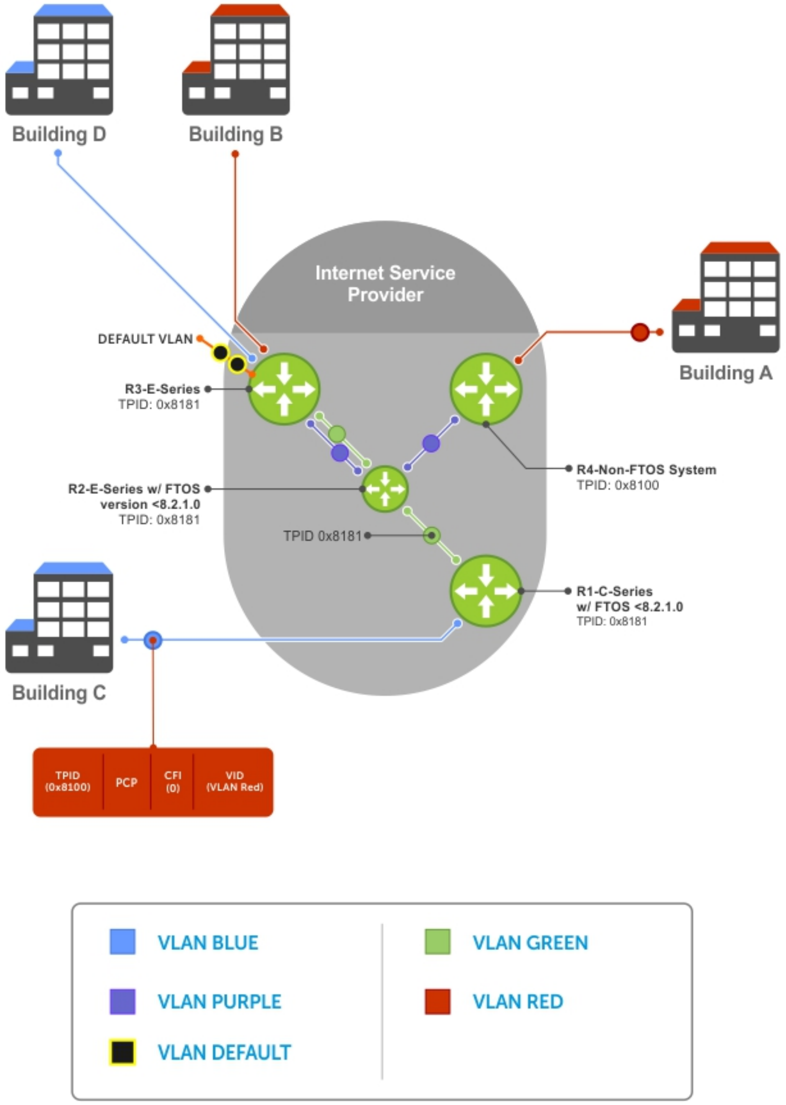

Building D
Building B
Internet Service Provider
Building A
DEFAULT VLAN
R3-E-Series
TPID: 0x8181
R4-Non-FTOS System
TPID: 0x8100
R2-E-Series w/ FTOS
version <8.2.1.0
TPID: 0x8181
TPID 0x8181
R1-C-Series
w/ FTOS <8.2.1.0
TPID: 0x8181
Building C
TPID (0x8100)
PCP
CFI (0)
VID (VLAN Red)
VLAN BLUE
VLAN PURPLE
VLAN DEFAULT
VLAN GREEN
VLAN RED

Figure 106. Single and Double-Tag First-byte TPID Match

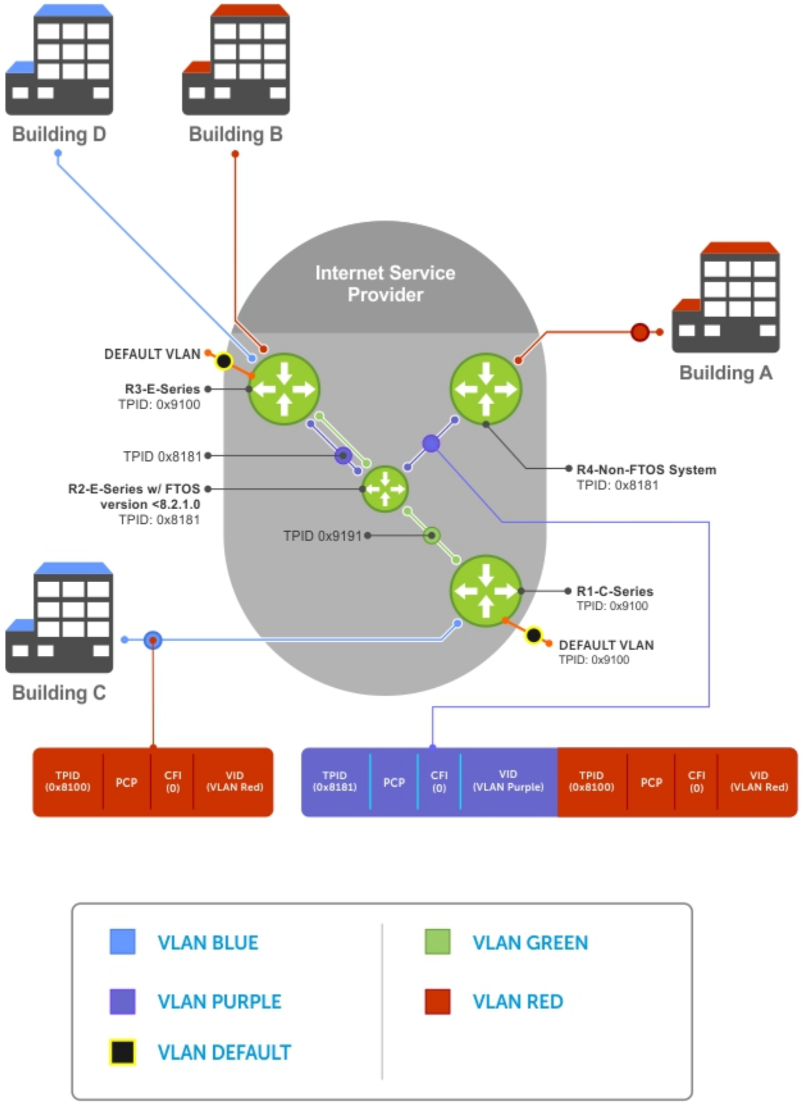

Building D
Building B
Internet Service Provider
Building A
DEFAULT VLAN
R3-E-Series
TPID: 0x9100
TPID 0x8181
R2-E-Series w/ FTOS version <8.2.1.0
TPID: 0x8181
TPID 0x9191
R4-Non-FTOS System
TPID: 0x8181
R1-C-Series
TPID: 0x9100
DEFAULT VLAN
TPID: 0x9100
Building C
TPID (0x8100)
PCP
CFI (0)
VID (VLAN Red)
TPID (0x8181)
PCP
CFI (0)
VID (VLAN Purple)
TPID (0x8100)
PCP
CFI (0)
VID (VLAN Red)
VLAN BLUE
VLAN PURPLE
VLAN DEFAULT
VLAN GREEN
VLAN RED

Figure 107. Single and Double-Tag TPID Mismatch

VLAN Stacking Packet Drop Precedence

VLAN stacking packet drop precedence is available on the S6000 platform.

The drop eligible indicator (DEI) bit in the S-Tag indicates to a service provider bridge which packets it should prefer to drop when congested.

Enabling Drop Eligibility

Enable drop eligibility globally before you can honor or mark the DEI value.
When you enable drop eligibility, DEI mapping or marking takes place according to the defaults. In this case, the CFI is affected according to the following table.

Table 64. Drop Eligibility Behavior

Ingress	Egress	DEI Disabled	DEI Enabled
Normal Port	Normal Port	Retain CFI	Set CFI to 0.
Trunk Port	Trunk Port	Retain inner tag CFI	Retain inner tag CFI.
		Retain outer tag CFI	Set outer tag CFI to 0.
Access Port	Trunk Port	Retain inner tag CFI	Retain inner tag CFI
		Set outer tag CFI to 0	Set outer tag CFI to 0

To enable drop eligibility globally, use the following command.

- Make packets eligible for dropping based on their DEI value.
  CONFIGURATION mode

  ```
  dei enable
  ```

  By default, packets are colored green, and DEI is marked 0 on egress.

Honoring the Incoming DEI Value

To honor the incoming DEI value, you must explicitly map the DEI bit to an Dell Networking OS drop precedence.
Precedence can have one of three colors.

Precedence	Description
Green	High-priority packets that are the least preferred to be dropped.
Yellow	Lower-priority packets that are treated as best-effort.
Red	Lowest-priority packets that are always dropped (regardless of congestion status).

- Honor the incoming DEI value by mapping it to an Dell Networking OS drop precedence.
  INTERFACE mode

  ```
  dei honor {0 | 1} {green | red | yellow}
  ```

  You may enter the command once for 0 and once for 1.

  Packets with an unmapped DEI value are colored green.

Example of Viewing DEI-Honoring Configuration

To display the DEI-honoring configuration, use the `show interface dei-honor [interface` *`slot/ port`* `| linecard` *`number`* `port-set` *`number`*`]` in EXEC Privilege mode.

```
Dell#show interface dei-honor

Default Drop precedence: Green
Interface CFI/DEI  Drop precedence
---------------------------------------
Gi 0/1    0        Green
Gi 0/1    1        Yellow
Gi 8/9    1        Red
Gi 8/40   0        Yellow
```

Marking Egress Packets with a DEI Value

On egress, you can set the DEI value according to a different mapping than ingress.
For ingress information, refer to Honoring the Incoming DEI Value.

To mark egress packets, use the following command.

- Set the DEI value on egress according to the color currently assigned to the packet.

  INTERFACE mode

  ```
  dei mark {green | yellow} {0 | 1}
  ```

Example of Viewing DEI-Marking Configuration

To display the DEI-marking configuration, use the `show interface dei-mark [`*`interface slot/port`*` | linecard `*`number`*` port-set `*`number`*`]` in EXEC Privilege mode.

```
Dell#show interface dei-mark

Default CFI/DEI Marking: 0
Interface Drop precedence CFI/DEI
--------------------------------
Gi 0/1    Green           0
Gi 0/1    Yellow          1
Gi 8/9    Yellow          0
Gi 8/40   Yellow          0
```

Dynamic Mode CoS for VLAN Stacking

One of the ways to ensure quality of service for customer VLAN-tagged frames is to use the 802.1p priority bits in the tag to indicate the level of QoS desired.

When an S-Tag is added to incoming customer frames, the 802.1p bits on the S-Tag may be configured statically for each customer or derived from the C-Tag using Dynamic Mode CoS. Dynamic Mode CoS maps the C-Tag 802.1p value to a S-Tag 802.1p value.

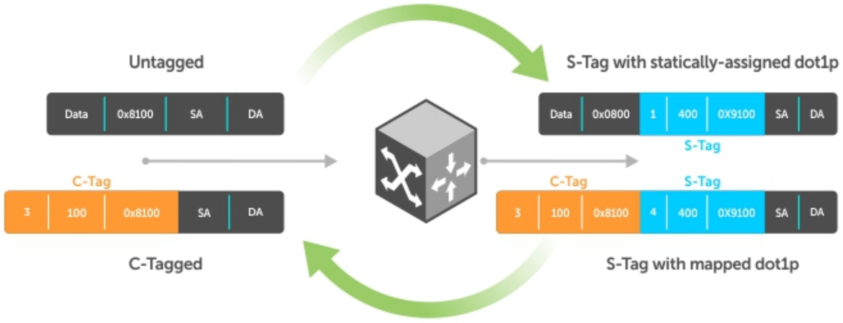

Untagged
Data
0x8100
SA
DA
S-Tag with statically-assigned dot1p
Data
0x0800
1
400
0X9100
SA
DA
S-Tag
C-Tag
3
100
0x8100
SA
DA
C-Tagged
C-Tag
S-Tag
3
100
0x8100
4
400
0X9100
SA
DA
S-Tag with mapped dot1p

Figure 108. Statically and Dynamically Assigned dot1p for VLAN Stacking

When configuring Dynamic Mode CoS, you have two options:

- Mark the S-Tag dot1p and queue the frame according to the original C-Tag dot1p. In this case, you must have other dot1p QoS configurations; this option is classic dot1p marking.
- Mark the S-Tag dot1p and queue the frame according to the S-Tag dot1p. For example, if frames with C-Tag dot1p values 0, 6, and 7 are mapped to an S-Tag dot1p value 0, all such frames are sent to the queue associated with the S-Tag 802.1p value 0. This option requires two different CAM entries, each in a different Layer 2 ACL FP block.

NOTE: The ability to map incoming C-Tag dot1p to any S-Tag dot1p requires installing up to eight entries in the Layer 2 QoS and Layer 2 ACL table for each configured customer VLAN. The scalability of this feature is limited by the impact of the 1:8 expansion in these content addressable memory (CAM) tables.

Dell Networking OS Behavior: For Option A shown in the previous illustration, when there is a conflict between the queue selected by Dynamic Mode CoS (vlan-stack dot1p-mapping) and a QoS configuration, the queue selected by Dynamic Mode CoS takes precedence. However, rate policing for the queue is determined by QoS configuration. For example, the following access-port configuration maps all traffic to Queue 0:

```
vlan-stack dot1p-mapping c-tag-dot1p 0-7 sp-tag-dot1p 1
```

However, if the following QoS configuration also exists on the interface, traffic is queued to Queue 0 but is policed at 40Mbps (`qos-policy-input` for queue 3) because class-map "a" of Queue 3 also matches the traffic. This is an expected behavior.

Examples of QoS Interface Configuration and Rate Policing

```
policy-map-input in layer2
service-queue 3 class-map a qos-policy 3
!
class-map match-any a layer2
match mac access-group a
!
mac access-list standard a
seq 5 permit any
!
qos-policy-input 3 layer2
rate-police 40
```

Likewise, in the following configuration, packets with dot1p priority 0–3 are marked as dot1p 7 in the outer tag and queued to Queue 3. Rate policing is according to `qos-policy-input` 3. All other packets will have outer dot1p 0 and hence are queued to Queue 1. They are therefore policed according to `qos-policy-input` 1.

```
policy-map-input in layer2
  service-queue 1 qos-policy 1
  service-queue 3 qos-policy 3
!
qos-policy-input 1 layer2
  rate-police 10
!
qos-policy-input 3 layer2
  rate-police 30
!
interface GigabitEthernet 0/21
  no ip address
  switchport
  vlan-stack access
  vlan-stack dot1p-mapping c-tag-dot1p 0-3 sp-tag-dot1p 7
```

```
service-policy input in layer2
no shutdown
```

Mapping C-Tag to S-Tag dot1p Values

To map C-Tag dot1p values to S-Tag dot1p values and mark the frames accordingly, use the following commands.

1. Allocate CAM space to enable queuing frames according to the C-Tag or the S-Tag.

   CONFIGURATION mode

   ```
   cam-acl l2acl number ipv4acl number ipv6acl number ipv4qos number l2qos
   number l2pt number ipmacacl number ecfmacl number {vman-qos | vman-qos-dual-
   fp} number
   ```

   - `vman-qos`: mark the S-Tag dot1p and queue the frame according to the original C-Tag dot1p. This method requires half as many CAM entries as `vman-qos-dual-fp`.
   - `vman-qos-dual-fp`: mark the S-Tag dot1p and queue the frame according to the S-Tag dot1p. This method requires twice as many CAM entries as `vman-qos` and FP blocks in multiples of 2.

   The default is: 0 FP blocks for `vman-qos` and `vman-qos-dual-fp`.
2. The new CAM configuration is stored in NVRAM and takes effect only after a save and reload.

   EXEC Privilege mode

   ```
   copy running-config startup-config reload
   ```
3. Map C-Tag dot1p values to a S-Tag dot1p value.

   INTERFACE mode

   ```
   vlan-stack dot1p-mapping c-tag-dot1p values sp-tag-dot1p value
   ```

   Separate C-Tag values by commas. Dashed ranges are permitted.

   Dynamic Mode CoS overrides any Layer 2 QoS configuration in case of conflicts.

NOTE: Because `dot1p-mapping` marks *and* queues packets, the only remaining applicable QoS configuration is rate metering. You may use Rate Shaping or Rate Policing.

Layer 2 Protocol Tunneling

Spanning tree bridge protocol data units (BPDUs) use a reserved destination MAC address called the bridge group address, which is 01-80-C2-00-00-00.

Only spanning-tree bridges on the local area network (LAN) recognize this address and process the BPDU. When you use VLAN stacking to connect physically separate regions of a network, BPDUs attempting to traverse the intermediate network might be consumed and later dropped because the intermediate network itself might be using spanning tree (shown in the following illustration).

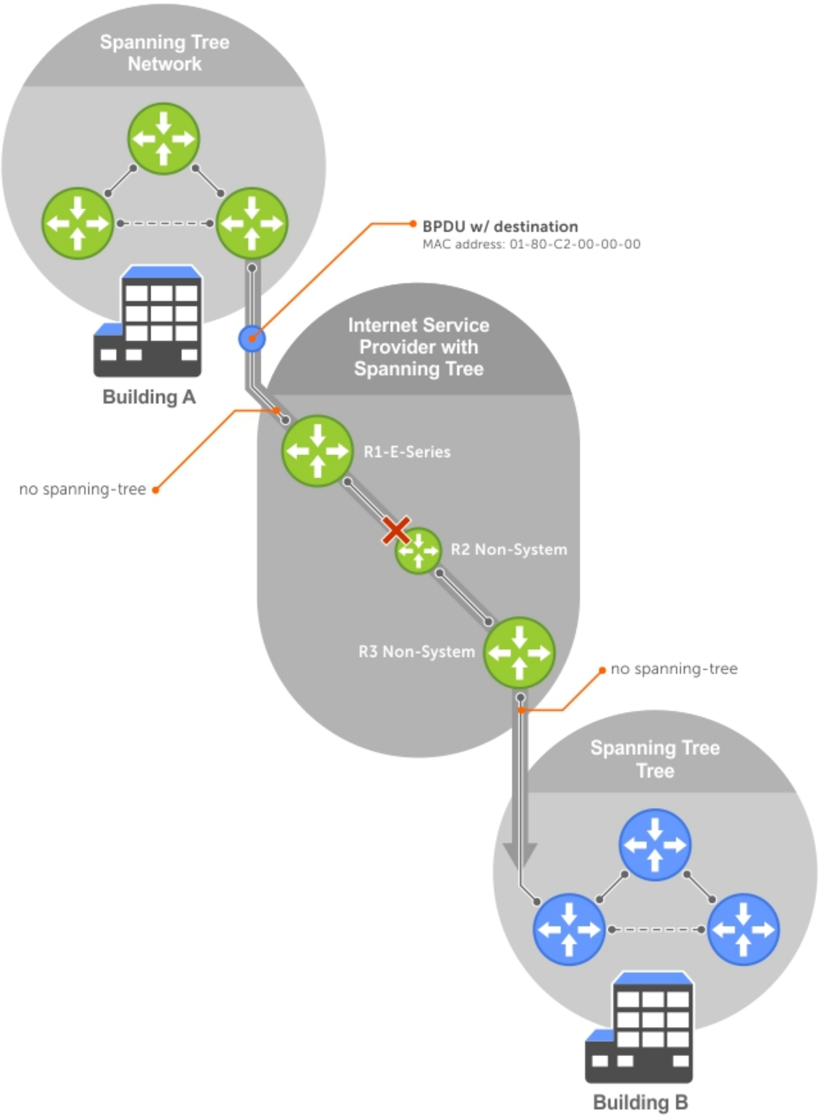

Spanning Tree Network
BPDU w/ destination
MAC address: 01-80-C2-00-00-00
Internet Service Provider with Spanning Tree
Building A
no spanning-tree
R1-E-Series
R2 Non-System
R3 Non-System
no spanning-tree
Spanning Tree Tree
Building B

Figure 109. VLAN Stacking without L2PT

You might need to transport control traffic transparently through the intermediate network to the other region. Layer 2 protocol tunneling enables BPDUs to traverse the intermediate network by identifying frames with the Bridge Group Address, rewriting the destination MAC to a user-configured non-reserved address, and forwarding the frames. Because the frames now use a unique MAC address, BPDUs are treated as normal data frames by the switches in the intermediate network core. On egress edge of the intermediate network, the MAC address rewritten to the original MAC address and forwarded to the opposing network region (shown in the following illustration).

Dell Networking OS Behavior: In Dell Networking OS versions prior to 8.2.1.0, the MAC address that Dell Networking systems use to overwrite the Bridge Group Address on ingress was non-configurable. The value of the L2PT MAC address was the Dell Networking-unique MAC address, 01-01-e8-00-00-00. As such, with these Dell Networking OS versions, Dell Networking systems are required at the egress edge of the intermediate network because only Dell Networking OS could recognize the significance of the destination MAC address and rewrite it to the original Bridge Group Address. In Dell Networking OS version 8.2.1.0 and later, the L2PT MAC address is user-configurable, so you can specify an address that non-Dell Networking systems can recognize and rewrite the address at egress edge.

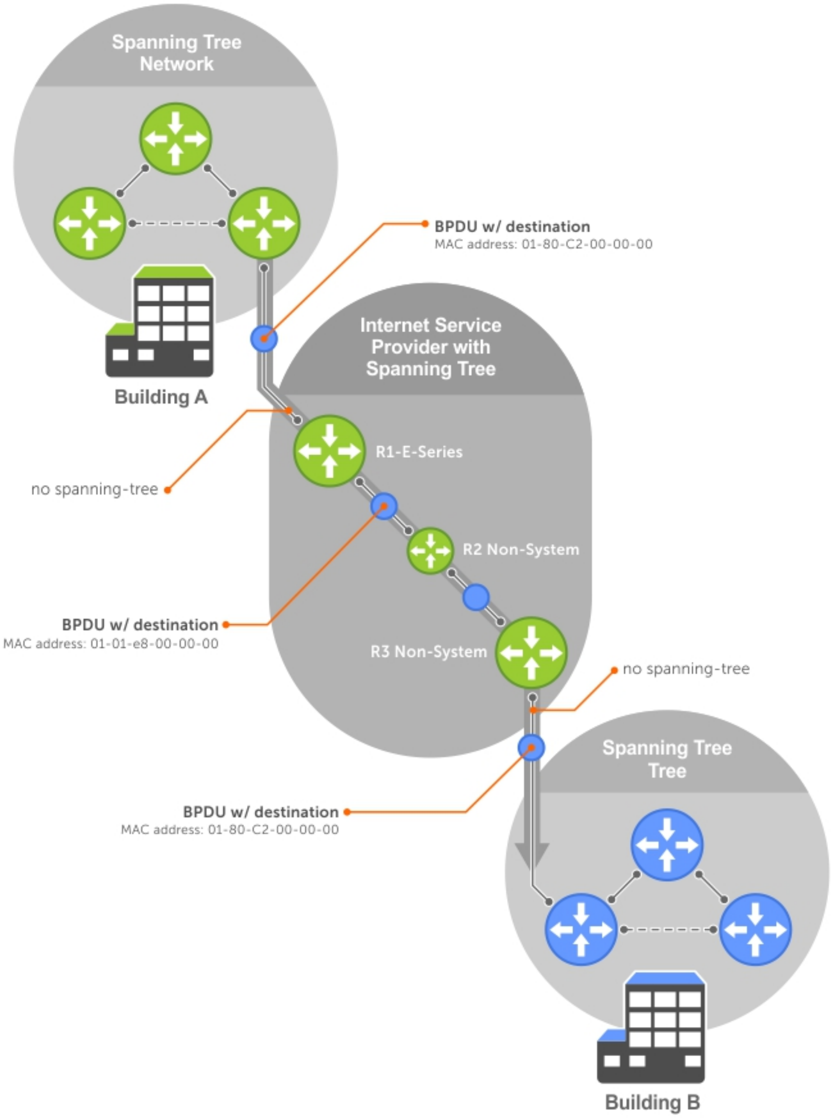

Spanning Tree Network
BPDU w/ destination
MAC address: 01-80-C2-00-00-00
Building A
Internet Service Provider with Spanning Tree
R1-E-Series
no spanning-tree
R2 Non-System
BPDU w/ destination
MAC address: 01-01-e8-00-00-00
R3 Non-System
no spanning-tree
Spanning Tree Tree
BPDU w/ destination
MAC address: 01-80-C2-00-00-00
Building B

Figure 110. VLAN Stacking with L2PT

Implementation Information

- L2PT is available for STP, RSTP, MSTP, and PVST+ BPDUs.
- No protocol packets are tunneled when you enable VLAN stacking.
- L2PT requires the default CAM profile.

Enabling Layer 2 Protocol Tunneling

To enable Layer 2 protocol tunneling, use the following command.

1. Verify that the system is running the default CAM profile. Use this CAM profile for L2PT.

   EXEC Privilege mode

   ```
   show cam-profile
   ```
2. Enable protocol tunneling globally on the system.

   CONFIGURATION mode

   ```
   protocol-tunnel enable
   ```
3. Tunnel BPDUs the VLAN.

   INTERFACE VLAN mode

   ```
   protocol-tunnel stp
   ```

Specifying a Destination MAC Address for BPDUs

By default, Dell Networking OS uses a Dell Networking-unique MAC address for tunneling BPDUs. You can configure another value.
To specify a destination MAC address for BPDUs, use the following command.

- Overwrite the BPDU with a user-specified destination MAC address when BPDUs are tunneled across the provider network.

  CONFIGURATION mode

  ```
  protocol-tunnel destination-mac
  ```

  The default is 01:01:e8:00:00:00

Setting Rate-Limit BPDUs

CAM space is allocated in sections called field processor (FP) blocks.
There are a total of 13 user-configurable FP blocks. The default number of blocks for L2PT is **0**; you must allocate at least one to enable BPDU rate-limiting.

To set the rate-lime BPDUs, use the following commands.

1. Create at least one FP group for L2PT.

   CONFIGURATION mode

   ```
   cam-acl l2acl
   ```

   For details about this command, refer to CAM Allocation.
2. Save the running-config to the startup-config.

   EXEC Privilege mode

   ```
   copy running-config startup-config
   ```
3. Reload the system.

   EXEC Privilege mode

   ```
   reload
   ```

4. Set a maximum rate at which the RPM processes BPDUs for L2PT.
   VLAN STACKING mode

   ```
   protocol-tunnel rate-limit
   ```

   The default is: no rate limiting.

   The range is from 64 to 320 kbps.

Debugging Layer 2 Protocol Tunneling

To debug Layer 2 protocol tunneling, use the following command.

- Display debugging information for L2PT.
  EXEC Privilege mode

  ```
  debug protocol-tunnel
  ```

Provider Backbone Bridging

IEEE 802.1ad—Provider Bridges amends 802.1Q—Virtual Bridged Local Area Networks so that service providers can use 802.1Q architecture to offer separate VLANs to customers with no coordination between customers, and minimal coordination between customers and the provider.

802.1ad specifies that provider bridges operating spanning tree use a reserved destination MAC address called the Provider Bridge Group Address, 01-80-C2-00-00-08, to exchange BPDUs instead of the Bridge Group Address, 01-80-C2-00-00-00, originally specified in 802.1Q. Only bridges in the service provider network use this destination MAC address so these bridges treat BPDUs originating from the customer network as normal data frames, rather than consuming them.

The same is true for GARP VLAN registration protocol (GVRP). 802.1ad specifies that provider bridges participating in GVRP use a reserved destination MAC address called the Provider Bridge GVRP Address, 01-80-C2-00-00-0D, to exchange GARP PDUs instead of the GVRP Address, 01-80-C2-00-00-21, specified in 802.1Q. Only bridges in the service provider network use this destination MAC address so these bridges treat GARP PDUs originating from the customer network as normal data frames, rather than consuming them.

Provider backbone bridging through IEEE 802.1ad eliminates the need for tunneling BPDUs with L2PT and increases the reliability of provider bridge networks as the network core need only learn the MAC addresses of core switches, as opposed to all MAC addresses received from attached customer devices.

43

sFlow

Configuring sFlow is supported on the S6000 platform.

Overview

The Dell Networking Operating System (OS) supports sFlow version 5.

sFlow is a standard-based sampling technology embedded within switches and routers which is used to monitor network traffic. It is designed to provide traffic monitoring for high-speed networks with many switches and routers. sFlow uses two types of sampling:

- Statistical packet-based sampling of switched or routed packet flows.
- Time-based sampling of interface counters.

The sFlow monitoring system consists of an sFlow agent (embedded in the switch/router) and an sFlow collector. The sFlow agent resides anywhere within the path of the packet and combines the flow samples and interface counters into sFlow datagrams and forwards them to the sFlow collector at regular intervals. The datagrams consist of information on, but not limited to, packet header, ingress and egress interfaces, sampling parameters, and interface counters.

Application-specific integrated circuits (ASICs) typically complete packet sampling. sFlow collector analyses the sFlow datagrams received from different devices and produces a network-wide view of traffic flows.

Implementation Information

Dell Networking sFlow is designed so that the hardware sampling rate is per line card port-pipe and is decided based on all the ports in that port-pipe.

If you do not enable sFlow on any port specifically, the global sampling rate is downloaded to that port and is to calculate the port-pipe’s lowest sampling rate. This design supports the possibility that sFlow might be configured on that port in the future. Back-off is triggered based on the port-pipe’s hardware sampling rate.

For example, if port 1 in the port-pipe has sFlow configured with a 16384 sampling rate while port 2 in the port-pipe has sFlow configured but no sampling rate set, Dell Networking OS applies a global sampling rate of 512 to port 2. The hardware sampling rate on the port-pipe is then set at 512 because that is the lowest configured rate on the port-pipe. When a high traffic situation occurs, a back-off is triggered and the hardware sampling rate is backed-off from 512 to 1024. Note that port 1 maintains its sampling rate of 16384; port 1 is unaffected because it maintains its configured sampling rate of 16484.

To avoid the back-off, either increase the global sampling rate or configure all the line card ports with the desired sampling rate even if some ports have no sFlow configured.

Important Points to Remember

- The Dell Networking OS implementation of the sFlow MIB supports sFlow configuration via snmpset.
- Dell Networking recommends the sFlow Collector be connected to the Dell Networking chassis through a line card port rather than the route processor module (RPM) management Ethernet port.
- Dell Networking OS exports all sFlow packets to the collector. A small sampling rate can equate to many exported packets. A backoff mechanism is automatically applied to reduce this amount. Some sampled packets may be dropped when the exported packet rate is high and the backoff mechanism is about to or is starting to take effect. The dropEvent counter, in the sFlow packet, is always zero.
- Community list and local preference fields are not filled in extended gateway element in the sFlow datagram.
- 802.1P source priority field is not filled in extended switch element in sFlow datagram.
- Only Destination and Destination Peer AS number are packed in the *dst-as-path* field in extended gateway element.
- If the packet being sampled is redirected using policy-based routing (PBR), the sFlow datagram may contain incorrect extended gateway/router information.
- The source virtual local area network (VLAN) field in the extended switch element is not packed in case of routed packet.
- The destination VLAN field in the extended switch element is not packed in a Multicast packet.

Enabling Extended sFlow

The S-Series platforms support `extended-switch` information processing only.
Extended sFlow packs additional information in the sFlow datagram depending on the type of sampled packet. You can enable the following options:

- `extended-switch` — 802.1Q VLAN ID and 802.1p priority information.
- `extended-router` — Next-hop and source and destination mask length.
- `extended-gateway` — Source and destination AS number and the BGP next-hop.

NOTE: The entire AS path is not included. BGP community-list and local preference information are not included. These fields are assigned default values and are not interpreted by the collector.

- Enable extended sFlow.

  ```
  sflow [extended-switch] [extended-router] [extended-gateway] enable
  ```

  By default packing of any of the extended information in the datagram is disabled.
- Confirm that extended information packing is enabled.

  ```
  show sflow
  ```

Examples of Verifying Extended sFlow

The bold line shows that extended sFlow settings are enabled on all three types.

```
Dell#show sflow
sFlow services are enabled
Global default sampling rate: 4096
Global default counter polling interval: 15
Global extended information enabled: gateway, router, switch
1 collectors configured
Collector IP addr: 10.10.10.3, Agent IP addr: 10.10.0.0, UDP port: 6343
77 UDP packets exported
0 UDP packets dropped
165 sFlow samples collected
```

```
69 sFlow samples dropped due to sub-sampling
Linecard 1 Port set 0 H/W sampling rate 8192
Gi 1/16: configured rate 8192, actual rate 8192, sub-sampling rate 1
Gi 1/17: configured rate 16384, actual rate 16384, sub-sampling rate 2
Linecard 3 Port set 1 H/W sampling rate 16384
Gi 3/40: configured rate 16384, actual rate 16384, sub-sampling rate 1
```

If you did not enable any extended information, the show output displays the following (shown in bold).

```
Dell#show sflow
sFlow services are disabled
Global default sampling rate: 32768
Global default counter polling interval: 20
Global extended information enabled: none
0 collectors configured
0 UDP packets exported
0 UDP packets dropped
0 sFlow samples collected
0 sFlow samples dropped due to sub-sampling
```

Enabling and Disabling sFlow on an Interface

By default, sFlow is disabled on all interfaces.
This CLI is supported on physical ports and link aggregation group (LAG) ports.

To enable sFlow on a specific interface, use the following command.

- Enable sFlow on an interface.

  INTERFACE mode

  ```
  [no] sflow enable
  ```

To disable sFlow on an interface, use the `no` version of this command.

sFlow Show Commands

Dell Networking OS includes the following sFlow display commands.

- Displaying Show sFlow Globally
- Displaying Show sFlow on an Interface
- Displaying Show sFlow on a Line Card

Displaying Show sFlow Global

To view sFlow statistics, use the following command.

- Display sFlow configuration information and statistics.

  EXEC mode

  ```
  show sflow
  ```

Example of Viewing sFlow Configuration (Global)

The first bold line indicates sFlow is globally enabled.

The second bold lines indicate sFlow is enabled on linecards Gi 1/16 and Gi 1/17

```
Dell#show sflow
sFlow services are enabled
Global default sampling rate: 32768
Global default counter polling interval: 20
1 collectors configured
Collector IP addr: 133.33.33.53, Agent IP addr: 133.33.33.116, UDP port: 6343
77 UDP packets exported
0 UDP packets dropped
165 sFlow samples collected
69 sFlow samples dropped due to sub-sampling

Linecard 1 Port set 0 H/W sampling rate 8192
   Gi 1/16: configured rate 8192, actual rate 8192, sub-sampling rate 1
 Gi 1/17: configured rate 16384, actual rate 16384, sub-sampling rate 2
```

Displaying Show sFlow on an Interface

To view sFlow information on a specific interface, use the following command.

- Display sFlow configuration information and statistics on a specific interface.

  EXEC mode

  `show sflow interface` *`interface-name`*

Examples of the sFlow `show` Commands

The following example shows the `show sflow interface` command.

```
Dell#show sflow interface gigabitethernet 1/16
Gi 1/16
Configured sampling rate         :8192
Actual sampling rate             :8192
Sub-sampling rate                :2
Counter polling interval         :15
Samples rcvd from h/w            :33
Samples dropped for sub-sampling :6
```

The following example shows the `show running-config interface` command.

```
Dell#show running-config interface gigabitethernet 1/16
!
interface GigabitEthernet 1/16
  no ip address
  mtu 9252
  ip mtu 9234
  switchport
  sflow enable
  sflow sample-rate 8192
  no shutdown
```

Displaying Show sFlow on a Stack-unit

To view sFlow statistics on a specified stack-unit, use the following command.

- Display sFlow configuration information and statistics on the specified interface.

  EXEC mode

  `show sflow stack—unit` *`slot-number`*

Example of Viewing sFlow Configuration (Line Card)

```
Dell#show sflow stack-unit 1
stack-unit 1
  Samples rcvd from h/w            :165
  Samples dropped for sub-sampling :69
  Total UDP packets exported       :77
  UDP packets exported via RPM     :77
  UDP packets dropped              :
```

Configuring Specify Collectors

The `sflow collector` command allows identification of sFlow collectors to which sFlow datagrams are forwarded.
You can specify up to two sFlow collectors. If you specify two collectors, the samples are sent to both.

- Identify sFlow collectors to which sFlow datagrams are forwarded.
  CONFIGURATION mode

  ```
  sflow collector ip-address agent-addr ip-address [number [max-datagram-size
  number] ] | [max-datagram-size number ]
  ```

  The default UDP port is **6343**.

  The default max-datagram-size is **1400**.

Changing the Polling Intervals

The `sflow polling-interval` command configures the polling interval for an interface in the maximum number of seconds between successive samples of counters sent to the collector.
This command changes the global default counter polling (20 seconds) interval. You can configure an interface to use a different polling interval.

To configure the polling intervals globally (in CONFIGURATION mode) or by interface (in INTERFACE mode), use the following command.

- Change the global default counter polling interval.
  CONFIGURATION mode or INTERFACE mode

  ```
  sflow polling-interval interval value
  ```

  - *`interval value`*: in seconds.

  The range is from 15 to 86400 seconds.

  The default is **20 seconds**.

Back-Off Mechanism

If the sampling rate for an interface is set to a very low value, the CPU can get overloaded with flow samples under high-traffic conditions.

In such a scenario, a binary back-off mechanism gets triggered, which doubles the sampling-rate (halves the number of samples per second) for all interfaces. The backoff mechanism continues to double the sampling-rate until the CPU condition is cleared. This is as per sFlow version 5 draft. After the back-off changes the sample-rate, you must manually change the sampling rate to the desired value.

As a result of back-off, the actual sampling-rate of an interface may differ from its configured sampling rate. You can view the actual sampling-rate of the interface and the configured sample-rate by using the `show sflow` command.

sFlow on LAG ports

When a physical port becomes a member of a LAG, it inherits the sFlow configuration from the LAG port.

Enabling Extended sFlow

The S-Series platforms support `extended-switch` information processing only.
Extended sFlow packs additional information in the sFlow datagram depending on the type of sampled packet. You can enable the following options:

- `extended-switch` — 802.1Q VLAN ID and 802.1p priority information.
- `extended-router` — Next-hop and source and destination mask length.
- `extended-gateway` — Source and destination AS number and the BGP next-hop.

NOTE: The entire AS path is not included. BGP community-list and local preference information are not included. These fields are assigned default values and are not interpreted by the collector.

- Enable extended sFlow.

  ```
  sflow [extended-switch] [extended-router] [extended-gateway] enable
  ```

  By default packing of any of the extended information in the datagram is disabled.
- Confirm that extended information packing is enabled.

  ```
  show sflow
  ```

Examples of Verifying Extended sFlow

The bold line shows that extended sFlow settings are enabled on all three types.

```
Dell#show sflow
sFlow services are enabled
Global default sampling rate: 4096
Global default counter polling interval: 15
Global extended information enabled: gateway, router, switch
1 collectors configured
Collector IP addr: 10.10.10.3, Agent IP addr: 10.10.0.0, UDP port: 6343
77 UDP packets exported
0 UDP packets dropped
165 sFlow samples collected
69 sFlow samples dropped due to sub-sampling
Linecard 1 Port set 0 H/W sampling rate 8192
Gi 1/16: configured rate 8192, actual rate 8192, sub-sampling rate 1
Gi 1/17: configured rate 16384, actual rate 16384, sub-sampling rate 2
Linecard 3 Port set 1 H/W sampling rate 16384
Gi 3/40: configured rate 16384, actual rate 16384, sub-sampling rate 1
```

If you did not enable any extended information, the show output displays the following (shown in bold).

```
Dell#show sflow
sFlow services are disabled
Global default sampling rate: 32768
Global default counter polling interval: 20
Global extended information enabled: none
0 collectors configured
```

```
0 UDP packets exported
0 UDP packets dropped
0 sFlow samples collected
0 sFlow samples dropped due to sub-sampling
```

Important Points to Remember

- To export extended-gateway data, BGP must learn the IP destination address.
- If the IP destination address is not learned via BGP the Dell Networking system does not export extended-gateway data.
- If the IP source address is learned via IGP, *srcAS* and *srcPeerAS* are zero.
- The *srcAS* and *srcPeerAS* might be zero even though the IP source address is learned via BGP. The c system packs the *srcAS* and *srcPeerAS* information only if the route is learned via BGP and it is reachable via the ingress interface of the packet.
- The sFlow sampling functionality is supported only for egress traffic and not for ingress traffic.

The previous points are summarized in following table.

Table 65. Extended Gateway Summary

IP SA	IP DA	srcAS and srcPeerAS	dstAS and dstPeerAS	Description
static/ connected/IGP	static/ connected/IGP	—	—	Extended gateway data is not exported because there is no AS information.
static/ connected/IGP	BGP	0	Exported	src_as and src_peer_as are zero because there is no AS information for IGP.
BGP	static/ connected/IGP	— Exported	— Exported	Prior to Dell Networking OS version 7.8.1.0, extended gateway data is not exported because IP DA is not learned via BGP. Version 7.8.1.0 allows extended gateway information in cases where the source and destination IP addresses are learned by different routing protocols, and for cases

IP SA	IP DA	srcAS and srcPeerAS	dstAS and dstPeerAS	Description
				where is source is reachable over ECMP.
BGP	BGP	Exported	Exported	Extended gateway data is packed.

44

Simple Network Management Protocol (SNMP)

Simple network management protocol (SNMP) is supported on the S6000 platform.

NOTE: On Dell Networking routers, standard and private SNMP management information bases (MIBs) are supported, including all *Get* and a limited number of *Set* operations (such as `set vlan` and `copy cmd`).

Protocol Overview

Network management stations use SNMP to retrieve or alter management data from network elements.

A datum of management information is called a *managed object*; the value of a managed object can be static or variable. Network elements store managed objects in a database called a *management information base* (MIB).

MIBs are hierarchically structured and use object identifiers to address managed objects, but managed objects also have a textual name called an *object descriptor*.

Implementation Information

The following describes SNMP implementation information.

- Dell Networking OS supports SNMP version 1 as defined by RFC 1155, 1157, and 1212, SNMP version 2c as defined by RFC 1901, and SNMP version 3 as defined by RFC 2571.
- Dell Networking OS supports up to 16 trap receivers.
- Dell Networking OS implementation of the sFlow MIB supports sFlow configuration via SNMP sets.
- SNMP traps for the spanning tree protocol (STP) and multiple spanning tree protocol (MSTP) state changes are based on BRIDGE MIB (RFC 1483) for STP and IEEE 802.1 *draft ruzin-mstp-mib-02* for MSTP.

SNMPv3 Compliance With FIPS

This functionality is supported on the S6000 platform.

SNMPv3 is compliant with the Federal information processing standard (FIPS) cryptography standard. The Advanced Encryption Standard (AES) Cipher Feedback (CFB) 128-bit encryption algorithm is in compliance with RFC 3826. SNMPv3 provides multiple authentication and privacy options for user configuration. A subset of these options are the FIPS-approved algorithms: HMAC-SHA1-96 for authentication and AES128-CFB for privacy. The other options are not FIPS-approved algorithms because of known security weaknesses. The AES128-CFB privacy option is supported and is compliant with RFC 3826.

The SNMPv3 feature also uses a FIPS-validated cryptographic module for all of its cryptographic operations when the system is configured with the `fips mode enable` command in Global

Configuration mode. When the FIPS mode is enabled on the system, SNMPv3 operates in a FIPS-compliant manner, and only the FIPS-approved algorithm options are available for SNMPv3 user configuration. When the FIPS mode is disabled on the system, all options are available for SNMPv3 user configuration.

The following table describes the authentication and privacy options that can be configured when the FIPS mode is enabled or disabled:

FIPS Mode	Privacy Options	Authentication Options
Disabled	des56 (DES56-CBC)	md5 (HMAC-MD5-96)
	aes128 (AES128-CFB)	sha (HMAC-SHA1-96)
Enabled	aes128 (AES128-CFB)	sha (HMAC-SHA1-96)

To enable security for SNMP packets transferred between the server and the client, you can use the `snmp-server user` *`username`* `group` *`groupname`* `3 auth` *`authentication-type auth-password`* `priv aes128` *`priv-password`* command to specify that AES-CFB 128 encryption algorithm needs to be used.

```
Dell(conf)#snmp-server user snmpguy snmpmon 3 auth sha AArt61wq priv aes128
jntRR59a
```

In this example, for a specified user and a group, the AES128-CFB algorithm, the authentication password to enable the server to receive packets from the host, and the privacy password to encode the message contents are configured.

SHA authentication needs to be used with the AES-CFB128 privacy algorithm only when FIPS is enabled because SHA is then the only available authentication level. If FIPS is disabled, you can use MD5 authentication in addition to SHA authentication with the AES-CFB128 privacy algorithm

You cannot modify the FIPS mode if SNMPv3 users are already configured and present in the system. An error message is displayed if you attempt to change the FIPS mode by using the `fips mode enable` command in Global Configuration mode. You can enable or disable FIPS mode only if SNMPv3 users are not previously set up. If previously configured users exist on the system, you must delete the existing users before you change the FIPS mode.

Keep the following points in mind when you configure the AES128-CFB algorithm for SNMPv3:

1. SNMPv3 authentication provides only the `sha` option when the FIPS mode is enabled.
2. SNMPv3 privacy provides only the `aes128 privacy` option when the FIPS mode is enabled.
3. If you attempt to enable or disable FIPS mode and if any SNMPv3 users are previously configured, an error message is displayed stating you must delete all of the SNMP users before changing the FIPS mode.
4. A message is logged indicating whether FIPS mode is enabled for SNMPv3. This message is generated only when the first SNMPv3 user is configured because you can modify the FIPS mode only when users are not previously configured. This log message is provided to assist your system security auditing procedures.

Configuration Task List for SNMP

Configuring SNMP version 1 or version 2 requires a single step.

NOTE: The configurations in this chapter use a UNIX environment with net-snmp version 5.4. This environment is only one of many RFC-compliant SNMP utilities you can use to manage your Dell Networking system using SNMP. Also, these configurations use SNMP version 2c.

- Creating a Community

Configuring SNMP version 3 requires configuring SNMP users in one of three methods. Refer to Setting Up User-Based Security (SNMPv3).

Related Configuration Tasks

- Managing Overload on Startup
- Reading Managed Object Values
- Writing Managed Object Values
- Subscribing to Managed Object Value Updates using SNMP
- Copying Configuration Files via SNMP
- Manage VLANs Using SNMP
- Enabling and Disabling a Port using SNMP
- Fetch Dynamic MAC Entries using SNMP
- Deriving Interface Indices
- Monitor Port-channels

Important Points to Remember

- Typically, 5-second timeout and 3-second retry values on an SNMP server are sufficient for both LAN and WAN applications. If you experience a timeout with these values, increase the timeout value to greater than 3 seconds, and increase the retry value to greater than 2 seconds on your SNMP server.
- User ACLs override group ACLs.

Set up SNMP

As previously stated, Dell Networking OS supports SNMP version 1 and version 2 that are community-based security models.

The primary difference between the two versions is that version 2 supports two additional protocol operations (*informs operation* and *snmpgetbulk query*) and one additional object (*counter64 object*).

SNMP version 3 (SNMPv3) is a user-based security model that provides password authentication for user security and encryption for data security and privacy. Three sets of configurations are available for SNMP read/write operations: no password or privacy, password privileges, password and privacy privileges.

You can configure a maximum of 16 users even if they are in different groups.

Creating a Community

For SNMPv1 and SNMPv2, create a community to enable the community-based security in Dell Networking OS.
The management station generates requests to either retrieve or alter the value of a management object and is called the *SNMP manager*. A network element that processes SNMP requests is called an *SNMP agent*. An *SNMP community* is a group of SNMP agents and managers that are allowed to interact. Communities are necessary to secure communication between SNMP managers and agents; SNMP agents do not respond to requests from management stations that are not part of the community.

Dell Networking OS enables SNMP automatically when you create an SNMP community and displays the following message. You must specify whether members of the community may only retrieve values (read), or retrieve and alter values (read-write).

```
22:31:23: %STKUNIT0-P:CP %SNMP-6-SNMP_WARM_START: Agent Initialized - SNMP
WARM_START.
```

To choose a name for the community you create, use the following command.

- Choose a name for the community.

  CONFIGURATION mode

  ```
  snmp-server community name {ro | rw}
  ```

Example of Creating an SNMP Community

To view your SNMP configuration, use the `show running-config snmp` command from EXEC Privilege mode.

```
Dell(conf)#snmp-server community my-snmp-community ro
22:31:23: %STKUNIT0-P:CP %SNMP-6-SNMP_WARM_START: Agent Initialized - SNMP
WARM_START.
Dell#show running-config snmp
!
snmp-server community mycommunity ro
```

Setting Up User-Based Security (SNMPv3)

When setting up SNMPv3, you can set users up with one of the following three types of configuration for SNMP read/write operations.
Users are typically associated to an SNMP group with permissions provided, such as OID view.

- **noauth** — no password or privacy. Select this option to set up a user with no password or privacy privileges. This setting is the basic configuration. Users must have a group and profile that do not require password privileges.
- **auth** — password privileges. Select this option to set up a user with password authentication.
- **priv** — password and privacy privileges. Select this option to set up a user with password and privacy privileges.

To set up user-based security (SNMPv3), use the following commands.

- Configure the user with view privileges only (no password or privacy privileges).

  CONFIGURATION mode

  ```
  snmp-server user name group-name 3 noauth
  ```
- Configure an SNMP group with view privileges only (no password or privacy privileges).

  CONFIGURATION mode

```
snmp-server group group-name 3 noauth auth read name write name
```

- Configure an SNMPv3 view.

  CONFIGURATION mode

  ```
  snmp-server view view-name oid-tree {included | excluded}
  ```

  NOTE: To give a user read and write view privileges, repeat this step for each privilege type.
- Configure the user with an authorization password (password privileges only).

  CONFIGURATION mode

  ```
  snmp-server user name group-name 3 noauth auth md5 auth-password
  ```
- Configure an SNMP group (password privileges only).

  CONFIGURATION mode

  ```
  snmp-server group groupname {oid-tree} auth read name write name
  ```
- Configure an SNMPv3 view.

  CONFIGURATION mode

  ```
  snmp-server view view-name 3 noauth {included | excluded}
  ```

  NOTE: To give a user read and write privileges, repeat this step for each privilege type.
- Configure an SNMP group (with password or privacy privileges).

  CONFIGURATION mode

  ```
  snmp-server group group-name {oid-tree} priv read name write name
  ```
- Configure the user with a secure authorization password and privacy password.

  CONFIGURATION mode

  ```
  snmp-server user name group-name {oid-tree} auth md5 auth-password priv des56
  priv password
  ```
- Configure an SNMPv3 view.

  CONFIGURATION mode

  ```
  snmp-server view view-name oid-tree {included | excluded}
  ```

Select a User-based Security Type

```
Dell(conf)#snmp-server host 1.1.1.1 traps {oid tree} version 3 ?
auth          Use the SNMPv3 authNoPriv Security Level
noauth        Use the SNMPv3 noAuthNoPriv Security Level
priv          Use the SNMPv3 authPriv Security Level
Dell(conf)#snmp-server host 1.1.1.1 traps {oid tree} version 3 noauth ?
WORD          SNMPv3 user name
```

Reading Managed Object Values

You may only retrieve (read) managed object values if your management station is a member of the same community as the SNMP agent.
Dell Networking supports RFC 4001, Textual Conventions for Internet Work Addresses that defines values representing a type of internet address. These values display for ipAddressTable objects using the `snmpwalk` command.

There are several UNIX SNMP commands that read data.

- Read the value of a single managed object.

  ```
  snmpget -v version -c community agent-ip {identifier.instance |
  descriptor.instance}
  ```

- Read the value of the managed object directly below the specified object.

  ```
  snmpgetnext -v version -c community agent-ip {identifier.instance |
  descriptor.instance}
  ```

- Read the value of many objects at once.

  ```
  snmpwalk -v version -c community agent-ip {identifier.instance |
  descriptor.instance}
  ```

Examples of Reading the Value of Managed Objects

In the following example, the value "4" displays in the OID before the IP address for IPv4. For an IPv6 IP address, a value of "16" displays.

```
> snmpget -v 2c -c mycommunity 10.11.131.161 sysUpTime.0
DISMAN-EVENT-MIB::sysUpTimeInstance = Timeticks: (32852616) 3 days, 19:15:26.16
> snmpget -v 2c -c mycommunity 10.11.131.161 .1.3.6.1.2.1.1.3.0
```

The following example shows reading the value of the next managed object.

```
> snmpgetnext -v 2c -c mycommunity 10.11.131.161 .1.3.6.1.2.1.1.3.0
SNMPv2-MIB::sysContact.0 = STRING:
> snmpgetnext -v 2c -c mycommunity 10.11.131.161 sysContact.0
```

The following example shows reading the value of the many managed objects at one time.

```
> snmpwalk -v 2c -c mycommunity 10.11.131.161 .1.3.6.1.2.1.1
SNMPv2-MIB::sysDescr.0 = STRING: Dell  Real Time Operating System Software
Dell Operating System Version: 1.0
Dell Application Software Version: E_MAIN4.9.4.0.0
Copyright (c) 1999-2014 by Dell
Build Time: Mon May 12 14:02:22 PDT 2008
SNMPv2-MIB::sysObjectID.0 = OID: SNMPv2-SMI::enterprises.6027.1.3.1
DISMAN-EVENT-MIB::sysUpTimeInstance = Timeticks: (32920954) 3 days, 19:26:49.54
SNMPv2-MIB::sysContact.0 = STRING:
```

Writing Managed Object Values

You may only alter (write) a managed object value if your management station is a member of the same community as the SNMP agent, and the object is writable.
Use the following command to write or write-over the value of a managed object.

- To write or write-over the value of a managed object.

  ```
  snmpset -v version -c community agent-ip {identifier.instance |
  descriptor.instance}syntax value
  ```

Example of Writing the Value of a Managed Object

```
> snmpset -v 2c -c mycommunity 10.11.131.161 sysName.0 s "R5"
SNMPv2-MIB::sysName.0 = STRING: R5
```

Configuring Contact and Location Information using SNMP

You may configure system contact and location information from the Dell Networking system or from the management station using SNMP.
To configure system contact and location information from the Dell Networking system and from the management station using SNMP, use the following commands.

- (From a Dell Networking system) Identify the system manager along with this person’s contact information (for example, an email address or phone number).

  CONFIGURATION mode

  ```
  snmp-server contact text
  ```

  You may use up to 55 characters.

  The default is **None**.
- (From a Dell Networking system) Identify the physical location of the system (for example, San Jose, 350 Holger Way, 1st floor lab, rack A1-1).

  CONFIGURATION mode

  ```
  snmp-server location text
  ```

  You may use up to 55 characters.

  The default is **None**.
- (From a management station) Identify the system manager along with this person’s contact information (for example, an email address or phone number).

  CONFIGURATION mode

  ```
  snmpset -v version -c community agent-ip sysContact.0 s “contact-info”
  ```

  You may use up to 55 characters.

  The default is **None**.
- (From a management station) Identify the physical location of the system (for example, San Jose, 350 Holger Way, 1st floor lab, rack A1-1).

  CONFIGURATION mode

  ```
  snmpset -v version -c community agent-ip sysLocation.0 s “location-info”
  ```

  You may use up to 55 characters.

  The default is **None**.

Subscribing to Managed Object Value Updates using SNMP

By default, the Dell Networking system displays some unsolicited SNMP messages (traps) upon certain events and conditions.
You can also configure the system to send the traps to a management station. Traps cannot be saved on the system.

Dell Networking OS supports the following three sets of traps:

- **RFC 1157-defined traps** — coldStart, warmStart, linkDown, linkUp, authenticationFailure, and egpNeighbborLoss.
- **Force10 enterpriseSpecific environment traps** — fan, supply, and temperature.
- **Force10 enterpriseSpecific protocol traps** — bgp, ecfm, stp, and xstp.

To configure the system to send SNMP notifications, use the following commands.

1. Configure the Dell Networking system to send notifications to an SNMP server.
   CONFIGURATION mode

   ```
   snmp-server host ip-address [traps | informs] [version 1 | 2c |3]
   [community-string]
   ```

   To send trap messages, enter the keyword `traps`.

   To send informational messages, enter the keyword `informs`.

   To send the SNMP version to use for notification messages, enter the keyword `version`.

   To identify the SNMPv1 community string, enter the name of the *`community-string`*.
2. Specify which traps the Dell Networking system sends to the trap receiver.
   CONFIGURATION mode

   ```
   snmp-server enable traps
   ```

   Enable all Dell Networking enterprise-specific and RFC-defined traps using the `snmp-server enable traps` command from CONFIGURATION mode.

   Enable all of the RFC-defined traps using the `snmp-server enable traps snmp` command from CONFIGURATION mode.

3. Specify the interfaces out of which Dell Networking OS sends SNMP traps.
   CONFIGURATION mode

   ```
   snmp-server trap-source
   ```

Example of RFC-Defined SNMP Traps and Related Enable Commands

The following example lists the RFC-defined SNMP traps and the command used to enable each. The *coldStart* and *warmStart* traps are enabled using a single command.

```
snmp authentication   SNMP_AUTH_FAIL:SNMP Authentication failed.Request with
invalid community string.
```

```
snmp coldstart        SNMP_COLD_START: Agent Initialized - SNMP COLD_START.
                      SNMP_WARM_START:Agent Initialized - SNMP WARM_START.
snmp linkdown         PORT_LINKDN:changed interface state to down:%d
snmp linkup           PORT_LINKUP:changed interface state to up:%d
```

Enabling a Subset of SNMP Traps

You can enable a subset of Dell Networking enterprise-specific SNMP traps using one of the following listed command options.
To enable a subset of Dell Networking enterprise-specific SNMP traps, use the following command.

- Enable a subset of SNMP traps.

  `snmp-server enable traps`

> NOTE: The `envmon` option enables all environment traps including those traps that are enabled with the `envmon supply`, `envmon temperature`, and `envmon fan` options.

Example of Dell Networking Enterprise-specific SNMP Traps

envmon

```
 STACK_STATE: Stack unit %d is in Active State
 STACKUNITUP: Stack unit 0 is up
```

envmon

```
  CARD_SHUTDOWN: %sLine card %d down - %s
  CARD_DOWN: %sLine card %d down - %s
  LINECARDUP: %sLine card %d is up
  CARD_MISMATCH: Mismatch: line card %d is type %s - type %s required.
  RPM_STATE: RPM1 is in Active State
  RPM_STATE: RPM0 is in Standby State
  RPM_DOWN: RPM 0 down - hard reset
  RPM_DOWN: RPM 0 down - card removed
  HOT_FAILOVER: RPM Failover Completed
  SFM_DISCOVERY: Found SFM 1
  SFM_REMOVE: Removed SFM 1
  MAJOR_SFM: Major alarm: Switch fabric down
  MAJOR_SFM_CLR: Major alarm cleared: Switch fabric up
  MINOR_SFM: MInor alarm: No working standby SFM
  MINOR_SFM_CLR: Minor alarm cleared: Working standby SFM present
  TASK SUSPENDED: SUSPENDED - svce:%d - inst:%d - task:%s
  RPM0-P:CP %CHMGR-2-CARD_PARITY_ERR
  ABNORMAL_TASK_TERMINATION: CRASH - task:%s %s
  CPU_THRESHOLD: Cpu %s usage above threshold. Cpu5SecUsage (%d)
  CPU_THRESHOLD_CLR: Cpu %s usage drops below threshold. Cpu5SecUsage (%d)
  MEM_THRESHOLD: Memory %s usage above threshold. MemUsage (%d)
  MEM_THRESHOLD_CLR: Memory %s usage drops below threshold. MemUsage (%d)
  DETECT_STN_MOVE: Station Move threshold exceeded for Mac %s in vlan %d
  CAM-UTILIZATION: Enable SNMP envmon CAM utilization traps.
```

envmon supply

```
  PEM_PRBLM: Major alarm: problem with power entry module %s
  PEM_OK: Major alarm cleared: power entry module %s is good
  MAJOR_PS: Major alarm: insufficient power %s
  MAJOR_PS_CLR: major alarm cleared: sufficient power
  MINOR_PS: Minor alarm: power supply non-redundant
  MINOR_PS_CLR: Minor alarm cleared: power supply redundant
```

envmon temperature

```
  MINOR_TEMP: Minor alarm: chassis temperature
  MINOR_TEMP_CLR: Minor alarm cleared: chassis temperature normal (%s %d
  temperature is within threshold of %dC)
  MAJOR_TEMP: Major alarm: chassis temperature high (%s temperature reaches or
  exceeds threshold of %dC)
  MAJOR_TEMP_CLR: Major alarm cleared: chassis temperature lower (%s %d
  temperature is within threshold of %dC)
```

envmon fan

```
  FAN_TRAY_BAD: Major alarm: fantray %d is missing or down
  FAN_TRAY_OK: Major alarm cleared: fan tray %d present
  FAN_BAD: Minor alarm: some fans in fan tray %d are down
  FAN_OK: Minor alarm cleared: all fans in fan tray %d are good
```

vlt

```
  Enable VLT traps.
```

vrrp

```
  Enable VRRP state change traps
```

xstp

```
  %SPANMGR-5-STP_NEW_ROOT: New Spanning Tree Root, Bridge ID Priority 32768,
  Address 0001.e801.fc35.
  %SPANMGR-5-STP_TOPOLOGY_CHANGE: Bridge port TeGigabitEthernet 11/38
transitioned
  from Forwarding to Blocking state.
  %SPANMGR-5-MSTP_NEW_ROOT_BRIDGE: Elected root bridge for instance 0.
  %SPANMGR-5-MSTP_NEW_ROOT_PORT: MSTP root changed to port TeGi 11/38 for
instance
  0. My Bridge ID: 40960:0001.e801.fc35 Old Root: 40960:0001.e801.fc35 New Root:
  32768:00d0.038a.2c01.
  %SPANMGR-5-MSTP_TOPOLOGY_CHANGE: Topology change BridgeAddr: 0001.e801.fc35
Mstp
  Instance Id 0 port TeGi 11/38 transitioned from forwarding to discarding
state.
```

ecfm

```
  %ECFM-5-ECFM_XCON_ALARM: Cross connect fault detected by MEP 1 in Domain
  customer1 at Level 7 VLAN 1000
  %ECFM-5-ECFM_ERROR_ALARM: Error CCM Defect detected by MEP 1 in Domain
customer1
  at Level 7 VLAN 1000
  %ECFM-5-ECFM_MAC_STATUS_ALARM: MAC Status Defect detected by MEP 1 in Domain
  provider at Level 4 VLAN 3000
  %ECFM-5-ECFM_REMOTE_ALARM: Remote CCM Defect detected by MEP 3 in Domain
  customer1 at Level 7 VLAN 1000
  %ECFM-5-ECFM_RDI_ALARM: RDI Defect detected by MEP 3 in Domain customer1 at
  Level 7 VLAN 1000
```

entity

```
  Enable entity change traps
  Trap SNMPv2-MIB::sysUpTime.0 = Timeticks: (1487406) 4:07:54.06,
  SNMPv2-MIB::snmpTrapOID.0 = OID: SNMPv2-SMI::mib-2.47.2.0.1,
  SNMPv2-SMI::enterprises.6027.3.6.1.1.2.0 = INTEGER: 4
  Trap SNMPv2-MIB::sysUpTime.0 = Timeticks: (1488564) 4:08:05.64,
  SNMPv2-MIB::snmpTrapOID.0 = OID: SNMPv2-SMI::mib-2.47.2.0.1,
  SNMPv2-SMI::enterprises.6027.3.6.1.1.2.0 = INTEGER: 5
  Trap SNMPv2-MIB::sysUpTime.0 = Timeticks: (1489064) 4:08:10.64,
  SNMPv2-MIB::snmpTrapOID.0 = OID: SNMPv2-SMI::mib-2.47.2.0.1,
  SNMPv2-SMI::enterprises.6027.3.6.1.1.2.0 = INTEGER: 6
  Trap SNMPv2-MIB::sysUpTime.0 = Timeticks: (1489568)
  4:08:15.68,SNMPv2-MIB::snmpTrapOID.0 = OID: SNMPv2-SMI::mib-2.47.2.0.1,
  SNMPv2-SMI::enterprises.6027.3.6.1.1.2.0 = INTEGER: 7
```

```
<cr>
  SNMP Copy Config Command Completed
  %RPM0-P:CP %SNMP-4-RMON_RISING_THRESHOLD:  STACKUNIT0 rising threshold alarm
```

```
from
  SNMP OID <oid>
  %RPM0-P:CP %SNMP-4-RMON_FALLING_THRESHOLD:  STACKUNIT0 falling threshold
alarm from
  SNMP OID <oid>
  %RPM0-P:CP %SNMP-4-RMON_HC_RISING_THRESHOLD:  STACKUNIT0 high-capacity rising
threshold
  alarm from SNMP OID <oid>
```

Copy Configuration Files Using SNMP

To do the following, use SNMP from a remote client.

- copy the running-config file to the startup-config file
- copy configuration files from the Dell Networking system to a server
- copy configuration files from a server to the Dell Networking system

You can perform all of these tasks using IPv4 or IPv6 addresses. The examples in this section use IPv4 addresses; however, you can substitute IPv6 addresses for the IPv4 addresses in all of the examples.

The following table lists the relevant MIBs for these functions are.

Table 66. MIB Objects for Copying Configuration Files via SNMP

MIB Object	OID	Object Values	Description
copySrcFileType	.1.3.6.1.4.1.6027.3.5.1.1.1.1.2	1 = Dell Networking OS file 2 = running-config 3 = startup-config	Specifies the type of file to copy from. The range is: • If copySrcFileType is running-config or startup-config, the default copySrcFileLocation is flash. • If copySrcFileType is a binary file, you must also specify copySrcFileLocation and copySrcFileName.
copySrcFileLocation	.1.3.6.1.4.1.6027.3.5.1.1.1.1.3	1 = flash 2 = slot0 3 = tftp 4 = ftp 5 = scp 6 = usbflash	Specifies the location of source file. • If copySrcFileLocation is FTP or SCP, you must specify copyServerAddress, copyUserName, and copyUserPassword.
copySrcFileName	.1.3.6.1.4.1.6027.3.5.1.1.1.1.4	Path (if the file is not in the current directory) and filename.	Specifies name of the file. • If copySourceFileType

MIB Object	OID	Object Values	Description
			is set to running-config or startup-config, copySrcFileName is not required.
copyDestFileType	.1.3.6.1.4.1.6027.3.5.1.1.1.1.5	1 = Dell Networking OS file 2 = running-config 3 = startup-config	Specifies the type of file to copy to. • If copySourceFileType is running-config or startup-config, the default copyDestFileLocation is flash. • If copyDestFileType is a binary, you must specify copyDestFileLocation and copyDestFileName.
copyDestFileLocation	.1.3.6.1.4.1.6027.3.5.1.1.1.1.6	1 = flash 2 = slot0 3 = tftp 4 = ftp 5 = scp	Specifies the location of destination file. • If copyDestFileLocation is FTP or SCP, you must specify copyServerAddress, copyUserName, and copyUserPassword.
copyDestFileName	.1.3.6.1.4.1.6027.3.5.1.1.1.1.7	Path (if the file is not in the default directory) and filename.	Specifies the name of destination file.
copyServerAddress	.1.3.6.1.4.1.6027.3.5.1.1.1.1.8	IP Address of the server.	The IP address of the server. • If you specify copyServerAddress, you must also specify copyUserName and copyUserPassword.
copyUserName	.1.3.6.1.4.1.6027.3.5.1.1.1.1.9	Username for the server.	Username for the FTP, TFTP, or SCP server. • If you specify copyUserName, you must also specify copyUserPassword.
copyUserPassword	.1.3.6.1.4.1.6027.3.5.1.1.1.1.10	Password for the server.	Password for the FTP, TFTP, or SCP server.

Copying a Configuration File

To copy a configuration file, use the following commands.

NOTE: In UNIX, enter the `snmpset` command for help using the following commands. Place the *f10-copy-config.mib* file in the directory from which you are executing the `snmpset` command or in the snmpset tool path.

1. Create an SNMP community string with read/write privileges.

   CONFIGURATION mode

   ```
   snmp-server community community-name rw
   ```
2. Copy the *f10-copy-config.mib* MIB from the Dell iSupport web page to the server to which you are copying the configuration file.
3. On the server, use the `snmpset` command as shown in the following example.

   ```
   snmpset -v snmp-version -c community-name -m mib_path/f10-copy-config.mib
   force10system-ip-address mib-object.index {i | a | s} object-value...
   ```

   - Every specified object must have an object value and must precede with the keyword i. Refer to the previous table.
   - *`index`* must be unique to all previously executed `snmpset` commands. If an index value has been used previously, a message like the following appears. In this case, increment the index value and enter the command again.

     ```
     Error in packet.
     Reason: notWritable (that object does not support modification)
     Failed object: FTOS-COPY-CONFIG-MIB::copySrcFileType.101
     ```
   - To complete the command, use as many MIB objects in the command as required by the MIB object descriptions shown in the previous table.

   NOTE: You can use the entire OID rather than the object name. Use the form: *`OID.index i object-value`*.

To view more information, use the following options in the `snmpset` command.

- `-c`: View the community, either public or private.
- `-m`: View the MIB files for the SNMP command.
- `-r`: Number of retries using the option
- `-t`: View the timeout.
- `-v`: View the SNMP version (either 1, 2, 2d, or 3).

The following examples show the `snmpset` command to copy a configuration. These examples assume that:

- the server OS is UNIX
- you are using SNMP version 2c
- the community name is public
- the file *f10-copy-config.mib* is in the current directory or in the snmpset tool path

Copying Configuration Files via SNMP

To copy the running-config to the startup-config from the UNIX machine, use the following command.

- Copy the running-config to the startup-config from the UNIX machine.

  `snmpset -v 2c -c public` *`force10system-ip-address`* `copySrcFileType.`*`index`* `i 2 copyDestFileType.`*`index`* `i 3`

Examples of Copying Configuration Files

The following examples show the command syntax using MIB object names and the same command using the object OIDs. In both cases, a unique index number follows the object.

The following example shows copying configuration files using MIB object names.

```
> snmpset -v 2c -r 0 -t 60 -c private -m ./f10-copy-config.mib 10.10.10.10
copySrcFileType.101
i 2 copyDestFileType.101 i 3
FTOS-COPY-CONFIG-MIB::copySrcFileType.101 = INTEGER: runningConfig(2)
FTOS-COPY-CONFIG-MIB::copyDestFileType.101 = INTEGER: startupConfig(3)
```

The following example shows copying configuration files using OIDs.

```
> snmpset -v 2c -c public -m ./f10-copy-config.mib 10.10.10.10
.1.3.6.1.4.1.6027.3.5.1.1.1.1.2.100 i 2 .1.3.6.1.4.1.6027.3.5.1.1.1.1.5.100 i 3
FTOS-COPY-CONFIG-MIB::copySrcFileType.100 = INTEGER: runningConfig(2)
FTOS-COPY-CONFIG-MIB::copyDestFileType.100 = INTEGER: startupConfig(3)
```

Copying the Startup-Config Files to the Running-Config

To copy the startup-config to the running-config from a UNIX machine, use the following command.

- Copy the startup-config to the running-config from a UNIX machine.

  `snmpset -c private -v 2c` *`force10system-ip-address`* `copySrcFileType.`*`index`* `i 3 copyDestFileType.`*`index`* `i 2`

Examples of Copying Configuration Files from a UNIX Machine

The following example shows how to copy configuration files from a UNIX machine using the object name.

```
> snmpset -c public -v 2c -m ./f10-copy-config.mib 10.11.131.162
copySrcFileType.7 i 3
copyDestFileType.7 i 2
FTOS-COPY-CONFIG-MIB::copySrcFileType.7 = INTEGER: runningConfig(3)
FTOS-COPY-CONFIG-MIB::copyDestFileType.7 = INTEGER: startupConfig(2)
```

The following example shows how to copy configuration files from a UNIX machine using OID.

```
>snmpset -c public -v 2c 10.11.131.162 .1.3.6.1.4.1.6027.3.5.1.1.1.1.2.8 i 3
.1.3.6.1.4.1.6027.3.5.1.1.1.1.5.8 i 2
SNMPv2-SMI::enterprises.6027.3.5.1.1.1.1.2.8 = INTEGER: 3
SNMPv2-SMI::enterprises.6027.3.5.1.1.1.1.5.8 = INTEGER: 2
```

Copying the Startup-Config Files to the Server via FTP

To copy the startup-config to the server via FTP from the UNIX machine, use the following command.

Copy the startup-config to the server via FTP from the UNIX machine.

`snmpset -v 2c -c public -m ./f10-copy-config.mib` *`force10system-ip-address`* `copySrcFileType.`*`index`* `i 2 copyDestFileName.`*`index`* `s` *`filepath/filename`* `copyDestFileLocation.`*`index`* `i 4 copyServerAddress.`*`index`* `a` *`server-ip-address`*

```
copyUserName.index s server-login-id copyUserPassword.index s server-login-
password
```

- precede `server-ip-address` by the keyword `a`.
- precede the values for copyUsername and copyUserPassword by the keyword `s`.

Example of Copying Configuration Files via FTP From a UNIX Machine

```
> snmpset -v 2c -c private -m ./f10-copy-config.mib 10.10.10.10 copySrcFileType.
110 i 2
copyDestFileName.110 s /home/startup-config copyDestFileLocation.110 i 4
copyServerAddress.110
a 11.11.11.11 copyUserName.110 s mylogin copyUserPassword.110 s mypass
FTOS-COPY-CONFIG-MIB::copySrcFileType.110 = INTEGER: runningConfig(2)
FTOS-COPY-CONFIG-MIB::copyDestFileName.110 = STRING: /home/startup-config
FTOS-COPY-CONFIG-MIB::copyDestFileLocation.110 = INTEGER: ftp(4)
FTOS-COPY-CONFIG-MIB::copyServerAddress.110 = IpAddress: 11.11.11.11
FTOS-COPY-CONFIG-MIB::copyUserName.110 = STRING: mylogin
FTOS-COPY-CONFIG-MIB::copyUserPassword.110 = STRING: mypass
```

Copying the Startup-Config Files to the Server via TFTP

To copy the startup-config to the server via TFTP from the UNIX machine, use the following command.

NOTE: Verify that the file exists and its permissions are set to 777. Specify the relative path to the TFTP root directory.

- Copy the startup-config to the server via TFTP from the UNIX machine.

  ```
  snmpset -v 2c -c public -m ./f10-copy-config.mib force10system-ip-address
  copySrcFileType.index i 3 copyDestFileType.index i 1 copyDestFileName.index s
  filepath/filename copyDestFileLocation.index i 3 copyServerAddress.index a
  server-ip-address
  ```

Example of Copying Configuration Files via TFTP From a UNIX Machine

```
.snmpset -v 2c -c private -m ./f10-copy-config.mib 10.10.10.10
copySrcFileType.4 i 3
copyDestFileType.4 i 1
copyDestFileLocation.4 i 3
copyDestFileName.4 s /home/myfilename
copyServerAddress.4 a 11.11.11.11
```

Copy a Binary File to the Startup-Configuration

To copy a binary file from the server to the startup-configuration on the Dell Networking system via FTP, use the following command.

- Copy a binary file from the server to the startup-configuration on the Dell Networking system via FTP.

  ```
  snmpset -v 2c -c public -m ./f10-copy-config.mib force10system-ip-address
  copySrcFileType.index i 1 copySrcFileLocation.index i 4 copySrcFileName.index
  s filepath/filename copyDestFileType.index i 3 copyServerAddress.index a
  server-ip-address copyUserName.index s server-login-id copyUserPassword.index
  s server-login-password
  ```

Example of Copying a Binary File From the Server to the Startup-Configuration via FTP

```
> snmpset -v 2c -c private -m ./f10-copy-config.mib 10.10.10.10 copySrcFileType.
10 i 1
copySrcFileLocation.10 i 4 copyDestFileType.10 i 3 copySrcFileName.10 s /home/
```

```
myfilename
copyServerAddress.10 a 172.16.1.56 copyUserName.10 s mylogin copyUserPassword.
10 s mypass
```

Additional MIB Objects to View Copy Statistics

Dell Networking provides more MIB objects to view copy statistics, as shown in the following table.

Table 67. Additional MIB Objects for Copying Configuration Files via SNMP

MIB Object	OID	Values	Description
copyState	.1.3.6.1.4.1.6027.3.5.1.1.1.1.11	1= running 2 = successful 3 = failed	Specifies the state of the copy operation.
copyTimeStarted	.1.3.6.1.4.1.6027.3.5.1.1.1.1.12	Time value	Specifies the point in the up-time clock that the copy operation started.
copyTimeCompleted	.1.3.6.1.4.1.6027.3.5.1.1.1.1.13	Time value	Specifies the point in the up-time clock that the copy operation completed.
copyFailCause	.1.3.6.1.4.1.6027.3.5.1.1.1.1.14	1 = bad filename 2 = copy in progress 3 = disk full 4 = file exists 5 = file not found 6 = timeout 7 = unknown	Specifies the reason the copy request failed.
copyEntryRowStatus	.1.3.6.1.4.1.6027.3.5.1.1.1.1.15	Row status	Specifies the state of the copy operation. Uses CreateAndGo when you are performing the copy. The state is set to *active* when the copy is completed.

Obtaining a Value for MIB Objects

To obtain a value for any of the MIB objects, use the following command.

- Get a copy-config MIB object value.

  `snmpset -v 2c -c public -m ./f10-copy-config.mib` *`force10system-ip-address`* `[`*`OID.index`* `|` *`mib-object.index`*`]`

`index`: the index value used in the `snmpset` command used to complete the copy operation.

NOTE: You can use the entire OID rather than the object name. Use the form: *`OID.index`*.

Examples of Getting MIB Object Values

The following examples show the `snmpget` command to obtain a MIB object value. These examples assume that:

- the server OS is UNIX
- you are using SNMP version 2c
- the community name is public
- the file f10-copy-config.mib is in the current directory

NOTE: In UNIX, enter the `snmpset` command for help using this command.

The following examples show the command syntax using MIB object names and the same command using the object OIDs. In both cases, the same index number used in the `snmpset` command follows the object.

The following command shows how to get a MIB object value using the object name.

```
> snmpget -v 2c -c private -m ./f10-copy-config.mib 10.11.131.140
copyTimeCompleted.110
FTOS-COPY-CONFIG-MIB::copyTimeCompleted.110 = Timeticks: (1179831) 3:16:38.31
```

The following command shows how to get a MIB object value using OID.

```
> snmpget -v 2c -c private 10.11.131.140 .1.3.6.1.4.1.6027.3.5.1.1.1.1.13.110
SNMPv2-SMI::enterprises.6027.3.5.1.1.1.1.13.110 = Timeticks: (1179831)
3:16:38.31
```

Manage VLANs using SNMP

The qBridgeMIB managed objects in Q-BRIDGE-MIB, defined in RFC 2674, allows you to use SNMP to manage VLANs.

Creating a VLAN

To create a VLAN, use the dot1qVlanStaticRowStatus object.
The snmpset operation shown in the following example creates VLAN 10 by specifying a value of 4 for instance 10 of the dot1qVlanStaticRowStatus object.
Example of Creating a VLAN using SNMP

```
> snmpset -v2c -c mycommunity 123.45.6.78 .1.3.6.1.2.1.17.7.1.4.3.1.5.10 i 4
SNMPv2-SMI::mib-2.17.7.1.4.3.1.5.10 = INTEGER: 4
```

Assigning a VLAN Alias

Write a character string to the dot1qVlanStaticName object to assign a name to a VLAN.
Example of Assigning a VLAN Alias using SNMP

```
[Unix system output]

> snmpset -v2c -c mycommunity 10.11.131.185 .
1.3.6.1.2.1.17.7.1.4.3.1.1.1107787786 s "My
```

```
VLAN"
SNMPv2-SMI::mib-2.17.7.1.4.3.1.1.1107787786 = STRING: "My VLAN"

[Dell system output]

Dell#show int vlan 10
Vlan 10 is down, line protocol is down
Vlan alias name is: My VLAN
Address is 00:01:e8:cc:cc:ce, Current address is 00:01:e8:cc:cc:ce
Interface index is 1107787786
Internet address is not set
MTU 1554 bytes, IP MTU 1500 bytes
LineSpeed auto
```

Displaying the Ports in a VLAN

Dell Networking OS identifies VLAN interfaces using an interface index number that is displayed in the output of the `show interface vlan` command.

Add Tagged and Untagged Ports to a VLAN

The value dot1qVlanStaticEgressPorts object is an array of all VLAN members.

The dot1qVlanStaticUntaggedPorts object is an array of only untagged VLAN members. All VLAN members that are not in dot1qVlanStaticUntaggedPorts are tagged.

- To add a tagged port to a VLAN, write the port to the dot1qVlanStaticEgressPorts object.
- To add an untagged port to a VLAN, write the port to the dot1qVlanStaticEgressPorts and dot1qVlanStaticUntaggedPorts objects.

NOTE: Whether adding a tagged or untagged port, specify values for both dot1qVlanStaticEgressPorts and dot1qVlanStaticUntaggedPorts.

Example of Adding an Untagged Port to a VLAN using SNMP

In the following example, Port 0/2 is added as an untagged member of VLAN 10.

```
>snmpset -v2c -c mycommunity 10.11.131.185 .
1.3.6.1.2.1.17.7.1.4.3.1.2.1107787786 x "40 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00"
.1.3.6.1.2.1.17.7.1.4.3.1.4.1107787786 x "40 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00
00 00 00 00 00 00 00 00 00"
SNMPv2-SMI::mib-2.17.7.1.4.3.1.2.1107787786 = Hex-STRING: 40 00 00 00 00 00 00
00 00 00 00
00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00
SNMPv2-SMI::mib-2.17.7.1.4.3.1.4.1107787786 = Hex-STRING: 40 00 00 00 00 00 00
00 00 00 00
00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00
```

Example of Adding a Tagged Port to a VLAN using SNMP

In the following example, Port 0/2 is added as a tagged member of VLAN 10.

```
>snmpset -v2c -c mycommunity 10.11.131.185 .
1.3.6.1.2.1.17.7.1.4.3.1.2.1107787786 x "40 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00"
.1.3.6.1.2.1.17.7.1.4.3.1.4.1107787786 x "00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00
00 00 00 00 00 00 00 00 00"
SNMPv2-SMI::mib-2.17.7.1.4.3.1.2.1107787786 = Hex-STRING: 40 00 00 00 00 00 00
00 00 00 00
00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00
SNMPv2-SMI::mib-2.17.7.1.4.3.1.4.1107787786 = Hex-STRING: 00 00 00 00 00 00 00
00 00 00 00
00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
```

Managing Overload on Startup

If you are running IS-IS, you can set a specific amount of time to prevent ingress traffic from being received after a reload and allow the routing protocol upgrade process to complete.
To prevent ingress traffic on a router while the IS reload is implemented, use the following command.

- Set the amount of time after an IS-IS reload is performed before ingress traffic is allowed at startup.

  `set-overload-bit on-startup isis`

The following OIDs are configurable through the `snmpset` command.

```
The node OID is 1.3.6.1.4.1.6027.3.18

F10-ISIS-MIB::f10IsisSysOloadSetOverload
F10-ISIS-MIB::f10IsisSysOloadSetOloadOnStartupUntil
F10-ISIS-MIB::f10IsisSysOloadWaitForBgp
F10-ISIS-MIB::f10IsisSysOloadV6SetOverload
F10-ISIS-MIB::f10IsisSysOloadV6SetOloadOnStartupUntil
F10-ISIS-MIB::f10IsisSysOloadV6WaitForBgp

    To enable overload bit for IPv4 set 1.3.6.1.4.1.6027.3.18.1.1 and IPv6 set
1.3.6.1.4.1.6027.3.18.1.4
    To set time to wait set 1.3.6.1.4.1.6027.3.18.1.2 and
1.3.6.1.4.1.6027.3.18.1.5
respectively
    To set time to wait till bgp session are up set 1.3.6.1.4.1.6027.3.18.1.3
and
1.3.6.1.4.1.6027.3.18.1.6
```

Enabling and Disabling a Port using SNMP

To enable and disable a port using SNMP, use the following commands.

1. Create an SNMP community on the Dell system.

CONFIGURATION mode

```
snmp-server community
```

2. From the Dell Networking system, identify the interface index of the port for which you want to change the admin status.

   EXEC Privilege mode

   ```
   show interface
   ```

   Or, from the management system, use the `snmpwwalk` command to identify the interface index.

3. Enter the `snmpset` command to change the admin status using either the object descriptor or the OID.

   ```
   snmpset with descriptor: snmpset -v version -c community agent-ip
   ifAdminStatus.ifindex i {1 | 2}

   snmpset with OID: snmpset -v version -c community agent-ip .
   1.3.6.1.2.1.2.2.1.7.ifindex i {1 | 2}
   ```

   Choose integer `1` to change the admin status to Up, or `2` to change the admin status to Down.

Fetch Dynamic MAC Entries using SNMP

Dell Networking supports the RFC 1493 dot1d table for the default VLAN and the dot1q table for all other VLANs.

NOTE: The 802.1q Q-BRIDGE MIB defines VLANs regarding 802.1d, as 802.1d itself does not define them. As a switchport must belong a VLAN (the default VLAN or a configured VLAN), all MAC address learned on a switchport are associated with a VLAN. For this reason, the Q-Bridge MIB is used for MAC address query. Moreover, specific to MAC address query, the MAC address indexes dot1dTpFdbTable only for a single forwarding database, while dot1qTpFdbTable has two indices — VLAN ID and MAC address — to allow for multiple forwarding databases and considering that the same MAC address is learned on multiple VLANs. The VLAN ID is added as the first index so that MAC addresses are read by the VLAN, sorted lexicographically. The MAC address is part of the OID instance, so in this case, lexicographic order is according to the most significant octet.

Table 68. MIB Objects for Fetching Dynamic MAC Entries in the Forwarding Database

MIB Object	OID	MIB	Description
dot1dTpFdbTable	.1.3.6.1.2.1.17.4.3	Q-BRIDGE MIB	List the learned unicast MAC addresses on the default VLAN.
dot1qTpFdbTable	.1.3.6.1.2.1.17.7.1.2. 2	Q-BRIDGE MIB	List the learned unicast MAC addresses on non-default VLANs.
dot3aCurAggFdb Table	.1.3.6.1.4.1.6027.3.2. 1.1.5	F10-LINK-AGGREGATION -MIB	List the learned MAC addresses of aggregated links (LAG).

In the following example, R1 has one dynamic MAC address, learned off of port TeGigabitEthernet 1/21, which a member of the default VLAN, VLAN 1. The SNMP walk returns the values for dot1dTpFdbAddress, dot1dTpFdbPort, and dot1dTpFdbStatus.

Each object is comprised of an OID concatenated with an instance number. In the case of these objects, the instance number is the decimal equivalent of the MAC address; derive the instance number by converting each hex pair to its decimal equivalent. For example, the decimal equivalent of E8 is 232, and so the instance number for MAC address 00:01:e8:06:95:ac is.0.1.232.6.149.172.

The value of dot1dTpFdbPort is the port number of the port off which the system learns the MAC address. In this case, of TeGigabitEthernet 1/21, the manager returns the integer 118.

Example of Fetching MAC Addresses Learned on the Default VLAN Using SNMP

```
----------------MAC Addresses on Force10 System------------------
R1_E600#show mac-address-table
VlanId   Mac Address      Type     Interface  State
1       00:01:e8:06:95:ac Dynamic  TeGi 1/21    Active
----------------Query from Management Station----------------------
>snmpwalk -v 2c -c techpubs 10.11.131.162 .1.3.6.1.2.1.17.4.3.1
SNMPv2-SMI::mib-2.17.4.3.1.1.0.1.232.6.149.172 = Hex-STRING: 00 01 E8 06 95 AC
```

Example of Fetching MAC Addresses Learned on a Non-default VLAN Using SNMP

In the following example, GigabitEthernet 1/21 is moved to VLAN 1000, a non-default VLAN. To fetch the MAC addresses learned on non-default VLANs, use the object dot1qTpFdbTable. The instance number is the VLAN number concatenated with the decimal conversion of the MAC address.

```
---------------MAC Addresses on Force10 System------------
R1_E600#show mac-address-table
VlanId  Mac Address        Type     Interface  State
1000    00:01:e8:06:95:ac  Dynamic  TeGi 1/21    Active
---------------Query from Management Station----------------
>snmpwalk -v 2c -c techpubs 10.11.131.162 .1.3.6.1.2.1.17.7.1.2.2.1
```

Example of Fetching MAC Addresses Learned on a Port-Channel Using SNMP

Use dot3aCurAggFdbTable to fetch the learned MAC address of a port-channel. The instance number is the decimal conversion of the MAC address concatenated with the port-channel number.

```
--------------MAC Addresses on Force10 System-------------------
R1_E600(conf)#do show mac-address-table
VlanId  Mac Address        Type     Interface  State
1000    00:01:e8:06:95:ac  Dynamic  Po 1       Active
-------------Query from Management Station----------------------
>snmpwalk -v 2c -c techpubs 10.11.131.162 .1.3.6.1.4.1.6027.3.2.1.1.5
SNMPv2-SMI::enterprises.6027.3.2.1.1.5.1.1.1000.0.1.232.6.149.172.1 = INTEGER:
1000
SNMPv2-SMI::enterprises.6027.3.2.1.1.5.1.2.1000.0.1.232.6.149.172.1 = Hex-
STRING: 00 01 E8
06 95 AC
SNMPv2-SMI::enterprises.6027.3.2.1.1.5.1.3.1000.0.1.232.6.149.172.1 = INTEGER: 1
SNMPv2-SMI::enterprises.6027.3.2.1.1.5.1.4.1000.0.1.232.6.149.172.1 = INTEGER: 1
```

Deriving Interface Indices

Dell Networking OS assigns an interface number to each (configured or unconfigured) physical and logical interface.

The interface index is a binary number with bits that indicate the slot number, port number, interface type, and card type of the interface. Dell Networking OS converts this binary index number to decimal, and displays it in the output of the `show interface` command.

Starting from the least significant bit (LSB):

- the first 14 bits represent the card type
- the next 4 bits represent the interface type
- the next 7 bits represent the port number
- the next 5 bits represent the slot number
- the next 1 bit is 0 for a physical interface and 1 for a logical interface
- the next 1 bit is unused

For example, the index 72925242 is 100010110001100000000111010 in binary. The binary interface index for TeGigabitEthernet 1/21 of a 48-port 10/100/1000Base-T line card with RJ-45 interface. Notice that the physical/logical bit and the final, unused bit are not given. The interface is physical, so represent this type of interface by a 0 bit, and the unused bit is always 0. These 2 bits are not given because they are the most significant bits, and leading zeros are often omitted.

To display the interface number, use the following command.

- Display the interface index number.

  EXEC Privilege mode

  ```
  show interface
  ```

Example of Deriving the Interface Index Number

To view the system image on Flash Partition A, use the chSysSwInPartitionAImgVers object or, to view the system image on Flash Partition B, use the chSysSwInPartitionBImgVers object.

Table 69. MIB Objects for Viewing the System Image on Flash Partitions

MIB Object	OID	Description	MIB
chSysSwInPartitionAImgVers	1.3.6.1.4.1.6027.3.10.1.2.8.1.11	List the version string of the system image in Flash Partition A.	Chassis MIB
chSysSwInPartitionBImgVers	1.3.6.1.4.1.6027.3.10.1.2.8.1.12	List the version string of the system image in Flash Partition B.	Chassis MIB

The system image can also be retrieved by performing an SNMP walk on the following OID: MIB Object is chSysSwModuleTable and the OID is 1.3.6.1.4.1.6027.3.10.1.2.8.

```
Dell#show interface Tegig 1/21
TenGigabitEthernet 1/21 is up, line protocol is up
```

Monitor Port-Channels

To check the status of a Layer 2 port-channel, use f10LinkAggMib (.1.3.6.1.4.1.6027.3.2). In the following example, Po 1 is a switchport and Po 2 is in Layer 3 mode.

Example of SNMP Trap for Monitored Port-Channels

```
[senthilnathan@lithium ~]$ snmpwalk -v 2c -c public 10.11.1.1 .
1.3.6.1.4.1.6027.3.2.1.1
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.1.1 = INTEGER: 1
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.1.2 = INTEGER: 2
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.2.1 = Hex-STRING: 00 01 E8 13 A5 C7
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.2.2 = Hex-STRING: 00 01 E8 13 A5 C8
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.3.1 = INTEGER: 1107755009
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.3.2 = INTEGER: 1107755010
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.4.1 = INTEGER: 1
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.4.2 = INTEGER: 1
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.5.1 = Hex-STRING: 00 00
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.5.2 = Hex-STRING: 00 00
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.6.1 = STRING: "Gi 5/84 " << Channel
member for Po1
SNMPv2-SMI::enterprises.6027.3.2.1.1.1.1.6.2 = STRING: "Gi 5/85 " << Channel
member for Po2
dot3aCommonAggFdbIndex
SNMPv2-SMI::enterprises.6027.3.2.1.1.6.1.1.1107755009.1 = INTEGER: 1107755009
dot3aCommonAggFdbVlanId
SNMPv2-SMI::enterprises.6027.3.2.1.1.6.1.2.1107755009.1 = INTEGER: 1
dot3aCommonAggFdbTagConfig
SNMPv2-SMI::enterprises.6027.3.2.1.1.6.1.3.1107755009.1 = INTEGER: 2 (Tagged 1 or
Untagged 2)
dot3aCommonAggFdbStatus
SNMPv2-SMI::enterprises.6027.3.2.1.1.6.1.4.1107755009.1 = INTEGER: 1 << Status
active, 2 – status inactive
```

Example of Viewing Status of Learned MAC Addresses

If we learn MAC addresses for the LAG, status is shown for those as well.

```
dot3aCurAggVlanId
SNMPv2-SMI::enterprises.6027.3.2.1.1.4.1.1.1.0.0.0.0.0.1.1 = INTEGER: 1
dot3aCurAggMacAddr
SNMPv2-SMI::enterprises.6027.3.2.1.1.4.1.2.1.0.0.0.0.0.1.1 = Hex-STRING: 00 00
00 00 00 01
dot3aCurAggIndex
SNMPv2-SMI::enterprises.6027.3.2.1.1.4.1.3.1.0.0.0.0.0.1.1 = INTEGER: 1
dot3aCurAggStatus
SNMPv2-SMI::enterprises.6027.3.2.1.1.4.1.4.1.0.0.0.0.0.1.1 = INTEGER: 1 << Status
active, 2 – status
inactive
```

Example of Viewing Changed Interface State for Monitored Ports

Layer 3 LAG does not include this support. SNMP trap works for the Layer 2 / Layer 3 / default mode LAG.

```
SNMPv2-MIB::sysUpTime.0 = Timeticks: (8500842) 23:36:48.42
SNMPv2-MIB::snmpTrapOID.0 = OID: IF-MIB::linkDown
IF-MIB::ifIndex.33865785 = INTEGER: 33865785
SNMPv2-SMI::enterprises.6027.3.1.1.4.1.2 = STRING: "OSTATE_DN: Changed
interface state to down: TeGi 0/0"
2010-02-10 14:22:39 10.16.130.4 [10.16.130.4]:
SNMPv2-MIB::sysUpTime.0 = Timeticks: (8500842) 23:36:48.42
SNMPv2-MIB::snmpTrapOID.0 = OID: IF-MIB::linkDown
IF-MIB::ifIndex.1107755009 = INTEGER: 1107755009
```

```
SNMPv2-SMI::enterprises.6027.3.1.1.4.1.2 = STRING: "OSTATE_DN: Changed interface
state to down: Po 1"
2010-02-10 14:22:40 10.16.130.4 [10.16.130.4]:
SNMPv2-MIB::sysUpTime.0 = Timeticks: (8500932) 23:36:49.32 SNMPv2-
MIB::snmpTrapOID.0 = OID:
IF-MIB::linkUp IF-MIB::ifIndex.33865785 = INTEGER: 33865785 SNMPv2-
SMI::enterprises.6027.3.1.1.4.1.2 =
STRING: "OSTATE_UP: Changed interface state to up: Gi 0/0"
2010-02-10 14:22:40 10.16.130.4 [10.16.130.4]:
SNMPv2-MIB::sysUpTime.0 = Timeticks: (8500934) 23:36:49.34 SNMPv2-
MIB::snmpTrapOID.0 = OID:
IF-MIB::linkUp IF-MIB::ifIndex.1107755009 = INTEGER: 1107755009
SNMPv2-SMI::enterprises.6027.3.1.1.4.1.2 = STRING: "OSTATE_UP: Changed interface
state to up: Po 1"
```

Troubleshooting SNMP Operation

When you use SNMP to retrieve management data from an SNMP agent on a Dell Networking router, take into account the following behavior.

- When you query an IPv4 icmpMsgStatsInPkts object in the ICMP table by using the `snmpwalk` command, the output for echo replies may be incorrectly displayed. To correctly display this information under ICMP statistics, use the `show ip traffic` command.
- When you query an icmpStatsInErrors object in the icmpStats table by using the `snmpget` or `snmpwalk` command, the output for IPv4 addresses may be incorrectly displayed. To correctly display this information under IP and ICMP statistics, use the `show ip traffic` command.
- When you query an IPv4 icmpMsgStatsInPkts object in the ICMP table by using the `snmpwalk` command, the echo response output may not be displayed. To correctly display ICMP statistics, such as echo response, use the `show ip traffic` command.

45

Storm Control

Storm control is supported on the S6000 platform.

The storm control feature allows you to control unknown-unicast and broadcast traffic on Layer 2 and Layer 3 physical interfaces.

Dell Networking Operating System (OS) Behavior: Dell Networking OS supports broadcast control (the `storm-control broadcast` command) for Layer 2 and Layer 3 traffic.

Configure Storm Control

Storm control is supported in INTERFACE mode and CONFIGURATION mode.

Configuring Storm Control from INTERFACE Mode

To configure storm control, use the following command.
From INTERFACE mode:

- You can only on configure storm control for ingress traffic.
- If you configure storm control from both INTERFACE and CONFIGURATION mode, the INTERFACE mode configurations override the CONFIGURATION mode configurations.
- The percentage of storm control is calculated based on the advertised rate of the line card, not by the speed setting.

- Configure storm control.

  INTERFACE mode

  ```
  storm control
  ```

Configuring Storm Control from CONFIGURATION Mode

To configure storm control from CONFIGURATION mode, use the following command.
From CONFIGURATION mode you can configure storm control for ingress and egress traffic.

Do not apply per-viritual local area network (VLAN) quality of service (QoS) on an interface that has storm-control enabled (either on an interface or globally).

- Configure storm control.

  CONFIGURATION mode

  ```
  storm control
  ```

46

Spanning Tree Protocol (STP)

The spanning tree protocol (STP) is supported on the S6000 platform.

Protocol Overview

STP is a Layer 2 protocol — specified by IEEE 802.1d — that eliminates loops in a bridged topology by enabling only a single path through the network.

By eliminating loops, the protocol improves scalability in a large network and allows you to implement redundant paths, which can be activated after the failure of active paths. Layer 2 loops, which can occur in a network due to poor network design and without enabling protocols like xSTP, can cause unnecessarily high switch CPU utilization and memory consumption.

Dell Networking OS supports three other variations of spanning tree, as shown in the following table.

Table 70. Dell Networking OS Supported Spanning Tree Protocols

Dell Networking Term	IEEE Specification
Spanning Tree Protocol (STP)	802.1d
Rapid Spanning Tree Protocol (RSTP)	802.1w
Multiple Spanning Tree Protocol (MSTP)	802.1s
Per-VLAN Spanning Tree Plus (PVST+)	Third Party

Configure Spanning Tree

Configuring spanning tree is a two-step process.

- Configuring Interfaces for Layer 2 Mode
- Enabling Spanning Tree Protocol Globally

Related Configuration Tasks

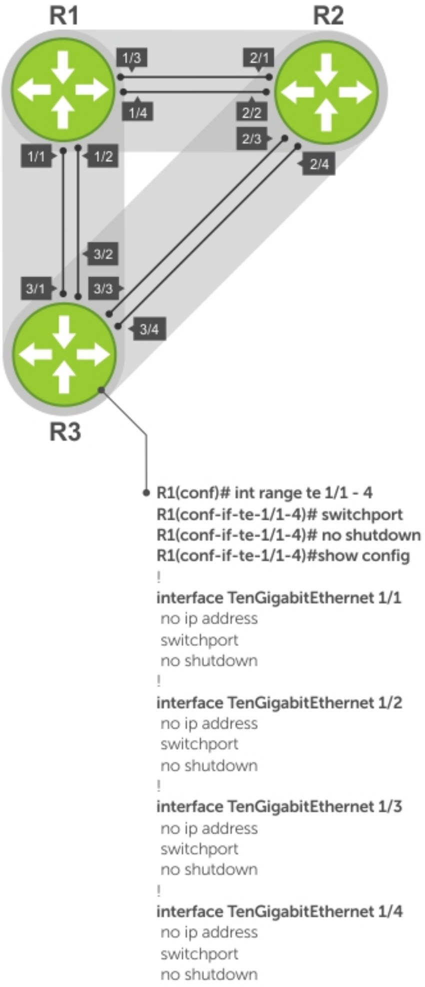
- Adding an Interface to the Spanning Tree Group
- Modifying Global Parameters
- Modifying Interface STP Parameters
- Enabling PortFast
- Prevent Network Disruptions with BPDU Guard
- STP Root Guard
- Enabling SNMP Traps for Root Elections and Topology Changes
- Configuring Spanning Trees as Hitless

Important Points to Remember

- STP is disabled by default.
- The Dell Networking OS supports only one spanning tree instance (0). For multiple instances, enable the multiple spanning tree protocol (MSTP) or per-VLAN spanning tree plus (PVST+). You may only enable one flavor of spanning tree at any one time.
- All ports in virtual local area networks (VLANs) and all enabled interfaces in Layer 2 mode are automatically added to the spanning tree topology at the time you enable the protocol.
- To add interfaces to the spanning tree topology after you enable STP, enable the port and configure it for Layer 2 using the `switchport` command.
- The IEEE Standard 802.1D allows 8 bits for port ID and 8 bits for priority. The 8 bits for port ID provide port IDs for 256 ports.

Configuring Interfaces for Layer 2 Mode

All interfaces on all switches that participate in spanning tree must be in Layer 2 mode and enabled.



```
R1(conf)# int range te 1/1 - 4
R1(conf-if-te-1/1-4)# switchport
R1(conf-if-te-1/1-4)# no shutdown
R1(conf-if-te-1/1-4)#show config
!
interface TenGigabitEthernet 1/1
 no ip address
 switchport
 no shutdown
!
interface TenGigabitEthernet 1/2
 no ip address
 switchport
 no shutdown
!
interface TenGigabitEthernet 1/3
 no ip address
 switchport
 no shutdown
!
interface TenGigabitEthernet 1/4
 no ip address
 switchport
 no shutdown
```

Figure 111. Example of Configuring Interfaces for Layer 2 Mode

To configure and enable the interfaces for Layer 2, use the following command.

1. If the interface has been assigned an IP address, remove it.
   INTERFACE mode

   ```
   no ip address
   ```
2. Place the interface in Layer 2 mode.
   INTERFACE

   ```
   switchport
   ```
3. Enable the interface.
   INTERFACE mode

   ```
   no shutdown
   ```

Example of the `show config` Command

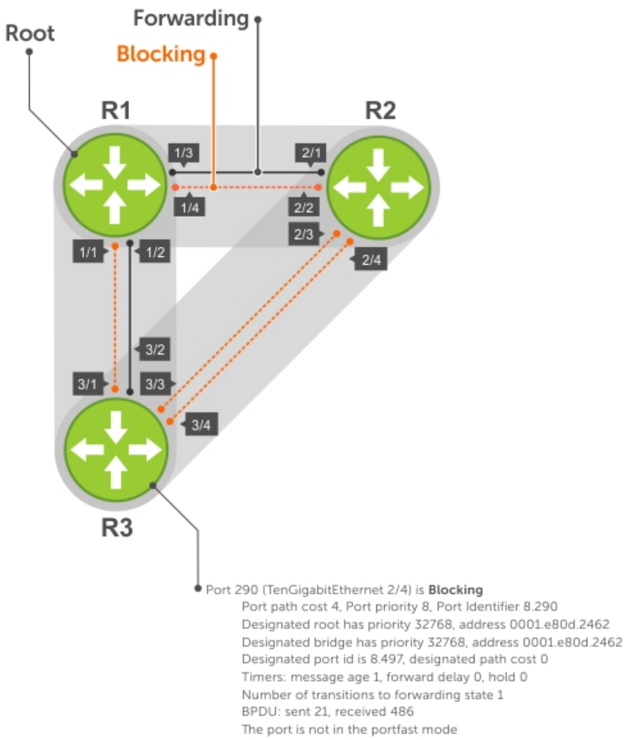
To verify that an interface is in Layer 2 mode and enabled, use the `show config` command from INTERFACE mode.

```
Dell(conf-if-gi-1/1)#show config
!
interface GigabitEthernet 1/1
  no ip address
  switchport
no shutdown
Dell(conf-if-gi-1/1)#
```

Enabling Spanning Tree Protocol Globally

Enable the spanning tree protocol globally; it is not enabled by default.
When you enable STP, all physical, VLAN, and port-channel interfaces that are enabled and in Layer 2 mode are automatically part of the Spanning Tree topology.

- Only one path from any bridge to any other bridge participating in STP is enabled.
- Bridges block a redundant path by disabling one of the link ports.



```
Port 290 (TenGigabitEthernet 2/4) is Blocking
    Port path cost 4, Port priority 8, Port Identifier 8.290
    Designated root has priority 32768, address 0001.e80d.2462
    Designated bridge has priority 32768, address 0001.e80d.2462
    Designated port id is 8.497, designated path cost 0
    Timers: message age 1, forward delay 0, hold 0
    Number of transitions to forwarding state 1
    BPDU: sent 21, received 486
    The port is not in the portfast mode
```

Figure 112. Spanning Tree Enabled Globally

To enable STP globally, use the following commands.

1. Enter PROTOCOL SPANNING TREE mode.
   CONFIGURATION mode
   ```
   protocol spanning-tree 0
   ```
2. Enable STP.
   PROTOCOL SPANNING TREE mode
   ```
   no disable
   ```

Examples of Verifying Spanning Tree Information

To disable STP globally for all Layer 2 interfaces, use the `disable` command from PROTOCOL SPANNING TREE mode.

To verify that STP is enabled, use the `show config` command from PROTOCOL SPANNING TREE mode.

```
Dell(conf)#protocol spanning-tree 0
Dell(config-span)#show config
!
protocol spanning-tree 0
  no disable
Dell#
```

To view the spanning tree configuration and the interfaces that are participating in STP, use the `show spanning-tree 0` command from EXEC privilege mode. If a physical interface is part of a port channel, only the port channel is listed in the command output.

```
R2#show spanning-tree 0
  Executing IEEE compatible Spanning Tree Protocol
    Bridge Identifier has priority 32768, address 0001.e826.ddb7
    Configured hello time 2, max age 20, forward delay 15
    Current root has priority 32768, address 0001.e80d.2462
    Root Port is 289 (GigabitEthernet 2/1), cost of root path is 4
    Topology change flag not set, detected flag not set
    Number of topology changes 3 last change occurred 0:16:11 ago
         from GigabitEthernet 2/3
    Timers: hold 1, topology change 35
            hello 2, max age 20, forward delay 15
    Times: hello 0, topology change 0, notification 0, aging Normal

  Port 289 (GigabitEthernet 2/1) is Forwarding
    Port path cost 4, Port priority 8, Port Identifier 8.289
    Designated root has priority 32768, address 0001.e80d.2462
    Designated bridge has priority 32768, address 0001.e80d.2462
    Designated port id is 8.496, designated path cost 0
    Timers: message age 1, forward delay 0, hold 0
    Number of transitions to forwarding state 1
    BPDU: sent 21, received 486
    The port is not in the portfast mode

  Port 290 (GigabitEthernet 2/2) is Blocking
    Port path cost 4, Port priority 8, Port Identifier 8.290
--More--
    Timers: message age 1, forward delay 0, hold 0
    Number of transitions to forwarding state 1
    BPDU: sent 21, received 486
    The port is not in the portfast mode
```

To confirm that a port is participating in Spanning Tree, use the `show spanning-tree 0 brief` command from EXEC privilege mode.

```
Dell#show spanning-tree 0 brief
  Executing IEEE compatible Spanning Tree Protocol
    Root ID Priority 32768, Address 0001.e80d.2462
    We are the root of the spanning tree
    Root Bridge hello time 2, max age 20, forward delay 15
    Bridge ID Priority 32768, Address 0001.e80d.2462
    Configured hello time 2, max age 20, forward delay 15
Interface                          Designated
Name    PortID Prio Cost Sts Cost  Bridge ID            PortID
-------------- ------ ---- ---- --- ----- --------------------
Gi 1/1  8.496  8     4 DIS   0     32768 0001.e80d.2462  8.496
Gi 1/2  8.497  8     4 DIS   0     32768 0001.e80d.2462  8.497
Gi 1/3  8.513  8     4 FWD   0     32768 0001.e80d.2462  8.513
Gi 1/4  8.514  8     4 FWD   0     32768 0001.e80d.2462  8.514
Dell#
```

Adding an Interface to the Spanning Tree Group

To add a Layer 2 interface to the spanning tree topology, use the following command.

- Enable spanning tree on a Layer 2 interface.
  INTERFACE mode

```
spanning-tree 0
```

Modifying Global Parameters

You can modify the spanning tree parameters. The root bridge sets the values for forward-delay, hello-time, and max-age and overwrites the values set on other bridges participating in STP.

NOTE: Dell Networking recommends that only experienced network administrators change the spanning tree parameters. Poorly planned modification of the spanning tree parameters can negatively affect network performance.

The following table displays the default values for STP.

Table 71. STP Default Values

STP Parameters	Default Value
Forward Delay	15 seconds
Hello Time	2 seconds
Max Age	20 seconds
Port Cost	
• 100-Mb/s Ethernet interfaces	• 19
• 1-Gigabit Ethernet interfaces	• 4
• 10-Gigabit Ethernet interfaces	• 2
• Port Channel with 100 Mb/s Ethernet interfaces	• 18
• Port Channel with 1-Gigabit Ethernet interfaces	• 3
• Port Channel with 10-Gigabit Ethernet interfaces	• 1
Port Priority	8

- Change the `forward-delay` parameter (the wait time before the interface enters the Forwarding state).

  PROTOCOL SPANNING TREE mode

  ```
  forward-delay seconds
  ```

  The range is from 4 to 30.

  The default is **15 seconds**.
- Change the `hello-time` parameter (the BPDU transmission interval).

  PROTOCOL SPANNING TREE mode

  ```
  hello-time seconds
  ```

  NOTE: With large configurations (especially those with more ports) Dell Networking recommends increasing the hello-time.

  The range is from 1 to 10.

  the default is **2 seconds**.
- Change the `max-age` parameter (the refresh interval for configuration information that is generated by recomputing the spanning tree topology).

PROTOCOL SPANNING TREE mode

```
max-age seconds
```

The range is from 6 to 40.

The default is **20 seconds**.

To view the current values for global parameters, use the `show spanning-tree 0` command from EXEC privilege mode. Refer to the second example in Enabling Spanning Tree Protocol Globally.

Modifying Interface STP Parameters

You can set the port cost and port priority values of interfaces in Layer 2 mode.

- **Port cost** — a value that is based on the interface type. The greater the port cost, the less likely the port is selected to be a forwarding port.
- **Port priority** — influences the likelihood that a port is selected to be a forwarding port in case that several ports have the same port cost.

The default values are listed in Modifying Global Parameters.

To change the port cost or priority of an interface, use the following commands.

- Change the port cost of an interface.
  INTERFACE mode

  ```
  spanning-tree 0 cost cost
  ```

  The range is from 0 to 65535.

  The default values are listed in Modifying Global Parameters.
- Change the port priority of an interface.
  INTERFACE mode

  ```
  spanning-tree 0 priority priority-value
  ```

  The range is from 0 to 15.

  The default is **8**.

To view the current values for interface parameters, use the `show spanning-tree 0` command from EXEC privilege mode. Refer to the second example in Enabling Spanning Tree Protocol Globally.

Enabling PortFast

The PortFast feature enables interfaces to begin forwarding traffic approximately 30 seconds sooner. Interfaces forward frames by default until they receive a BPDU that indicates that they should behave otherwise; they do not go through the Learning and Listening states. The `bpduguard shutdown-on-violation` option causes the interface hardware to be shut down when it receives a BPDU. When you only implement `bpduguard`, although the interface is placed in an Error Disabled state when receiving the BPDU, the physical interface remains up and spanning-tree drops packets in the hardware after a BPDU violation. BPDUs are dropped in the software after receiving the BPDU violation.

CAUTION: Enable PortFast only on links connecting to an end station. PortFast can cause loops if it is enabled on an interface connected to a network.

To enable PortFast on an interface, use the following command.

- Enable PortFast on an interface.
  INTERFACE mode

  ```
  spanning-tree stp-id portfast [bpduguard | [shutdown-on-violation]]
  ```

Example of Verifying PortFast is Enabled on an Interface

To verify that PortFast is enabled on a port, use the `show spanning-tree` command from EXEC Privilege mode or the `show config` command from INTERFACE mode. Dell Networking recommends using the `show config` command.

```
Dell#(conf-if-gi-1/1)#show conf
!
interface GigabitEthernet 1/1
  no ip address
  switchport
  spanning-tree 0 portfast
  no shutdown
Dell#(conf-if-gi-1/1)#
```

Prevent Network Disruptions with BPDU Guard

Configure the Portfast (and Edgeport, in the case of RSTP, PVST+, and MSTP) feature on ports that connect to end stations. End stations do not generate BPDUs, so ports configured with Portfast/ Edgport (edgeports) do not expect to receive BDPUs.

If an edgeport does receive a BPDU, it likely means that it is connected to another part of the network, which can negatively affect the STP topology. The BPDU Guard feature blocks an edgeport after receiving a BPDU to prevent network disruptions, and Dell Networking OS displays the following message.

```
3w3d0h: %RPM0-P:RP2 %SPANMGR-5-BPDU_GUARD_RX_ERROR: Received Spanning Tree BPDU
on
BPDU guard port. Disable GigabitEthernet 3/41.
```

Enable BPDU Guard using the `bpduguard` option when enabling PortFast or EdgePort. The `bpduguard shutdown-on-violation` option causes the interface hardware to be shut down when it receives a BPDU. Otherwise, although the interface is placed in an Error Disabled state when receiving the BPDU, the physical interface remains up and spanning-tree will only drop packets after a BPDU violation.

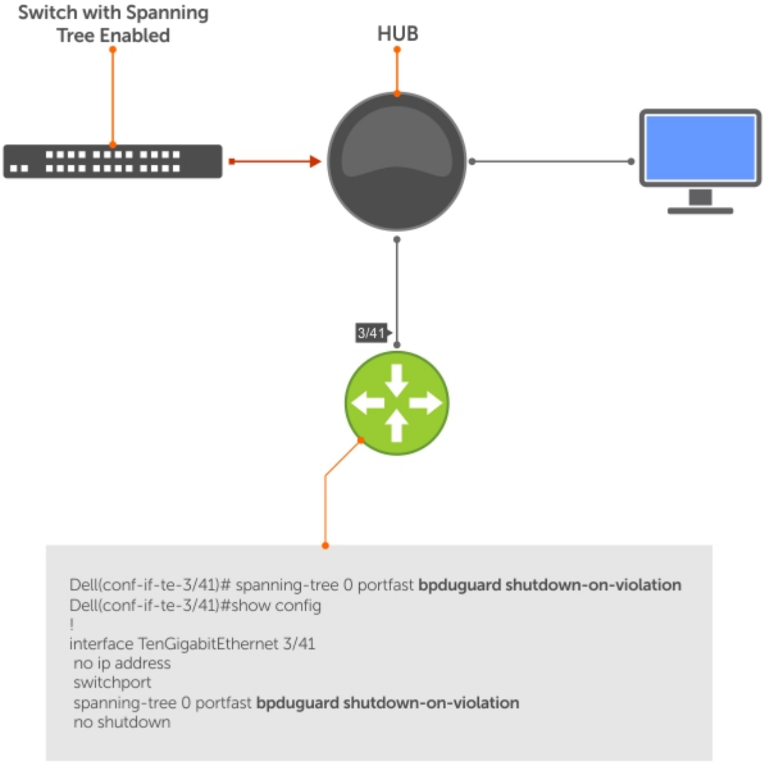
The following example shows a scenario in which an edgeport might unintentionally receive a BPDU. The port on the Dell Networking system is configured with Portfast. If the switch is connected to the hub, the BPDUs that the switch generates might trigger an undesirable topology change. If you enable BPDU Guard, when the edge port receives the BPDU, the BPDU is dropped, the port is blocked, and a console message is generated.

NOTE: Unless you enable the `shutdown-on-violation` option, spanning-tree only drops packets after a BPDU violation; the physical interface remains up.

Dell Networking OS Behavior: Regarding `bpduguard shutdown-on-violation` behavior:

- If the interface to be shut down is a port channel, all the member ports are disabled in the hardware.

- When you add a physical port to a port channel already in the Error Disable state, the new member port is also disabled in the hardware.
- When you remove a physical port from a port channel in the Error Disable state, the Error Disabled state is cleared on this physical port (the physical port is enabled in the hardware).
- The `reset linecard` command does not clear the Error Disabled state of the port or the Hardware Disabled state. The interface continues to be disables in the hardware.
- You can clear the Error Disabled state with any of the following methods:
  - Perform a `shutdown` command on the interface.
  - Disable the `shutdown-on-violation` command on the interface (the `no spanning-tree` *`stp-id`* `portfast [bpduguard | [shutdown-on-violation]]` command).
  - Disable spanning tree on the interface (the `no spanning-tree` command in INTERFACE mode).
  - Disabling global spanning tree (the `no spanning-tree` in CONFIGURATION mode).



```
Dell(conf-if-te-3/41)# spanning-tree 0 portfast bpduguard shutdown-on-violation
Dell(conf-if-te-3/41)#show config
!
interface TenGigabitEthernet 3/41
 no ip address
 switchport
 spanning-tree 0 portfast bpduguard shutdown-on-violation
 no shutdown
```

Figure 113. Enabling BPDU Guard

Dell Networking OS Behavior: BPDU guard and BPDU filtering both block BPDUs, but are two separate features.

BPDU guard:

- is used on edgeports and blocks all traffic on edgeport if it receives a BPDU.
- drops the BPDU after it reaches the RPM and generates a console message.

BPDU filtering:

- disables spanning tree on an interface
- drops all BPDUs at the line card without generating a console message

Example of Blocked BPDUs

```
Dell(conf-if-gi-0/7)#do show spanning-tree rstp brief
Executing IEEE compatible Spanning Tree Protocol
Root ID Priority 32768, Address 0001.e805.fb07
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID Priority 32768, Address 0001.e85d.0e90
Configured hello time 2, max age 20, forward delay 15

Interface                           Designated
Name    PortID   Prio Cost   Sts Cost  Bridge ID            PortID
---------- -------- ---- ------- --- ------- --------------------
Gi 0/6 128.263 128   20000 FWD 20000 32768 0001.e805.fb07 128.653
Gi 0/7 128.264 128   20000 EDS 20000 32768 0001.e85d.0e90 128.264

Interface
Name    Role    PortID   Prio Cost   Sts Cost   Link-type Edge
---------- ------ -------- ---- ------- --- ----------------
Gi 0/6 Root   128.263 128   20000 FWD 20000 P2P       No
Gi 0/7 ErrDis 128.264 128  20000 EDS 20000 P2P       No
Dell(conf-if-gi-0/7)#do show ip int br gi 0/7
Interface           IP-Address OK Method  Status Protocol
GigabitEthernet 0/7 unassigned YES Manual up     up
```

Selecting STP Root

The STP determines the root bridge, but you can assign one bridge a lower priority to increase the likelihood that it becomes the root bridge. You can also specify that a bridge is the root or the secondary root.
To change the bridge priority or specify that a bridge is the root or secondary root, use the following command.

- Assign a number as the bridge priority or designate it as the root or secondary root.

  PROTOCOL SPANNING TREE mode

  ```
  bridge-priority {priority-value | primary | secondary}
  ```

  - *`priority-value`*: the range is from 0 to 65535. The lower the number assigned, the more likely this bridge becomes the root bridge.

  The primary option specifies a bridge priority of 8192.

  The secondary option specifies a bridge priority of 16384.

  The default is **32768**.

Example of Viewing STP Root Information

To view only the root information, use the `show spanning-tree root` command from EXEC privilege mode.

```
Dell#show spanning-tree 0 root
  Root ID Priority 32768, Address 0001.e80d.2462
  We are the root of the spanning tree
```

```
  Root Bridge hello time 2, max age 20, forward delay 15
Dell#
```

STP Root Guard

STP root guard is supported on the platform.

Use the STP root guard feature in a Layer 2 network to avoid bridging loops. In STP, the switch in the network with the lowest priority (as determined by STP or set with the `bridge-priority` command) is selected as the root bridge. If two switches have the same priority, the switch with the lower MAC address is selected as the root. All other switches in the network use the root bridge as the reference used to calculate the shortest forwarding path.

Because any switch in an STP network with a lower priority can become the root bridge, the forwarding topology may not be stable. The location of the root bridge can change, resulting in unpredictable network behavior. The STP root guard feature ensures that the position of the root bridge does not change.

Root Guard Scenario

For example, as shown in the following illustration (STP topology 1, upper left) Switch A is the root bridge in the network core. Switch C functions as an access switch connected to an external device. The link between Switch C and Switch B is in a Blocking state. The flow of STP BPDUs is shown in the illustration.

In STP topology 2 (shown in the upper right), STP is enabled on device D on which a software bridge application is started to connect to the network. Because the priority of the bridge in device D is lower than the root bridge in Switch A, device D is elected as root, causing the link between Switches A and B to enter a Blocking state. Network traffic then begins to flow in the directions indicated by the BPDU arrows in the topology. If the links between Switches C and A or Switches C and B cannot handle the increased traffic flow, frames may be dropped.

In STP topology 3 (shown in the lower middle), if you have enabled the root guard feature on the STP port on Switch C that connects to device D, and device D sends a superior BPDU that would trigger the election of device D as the new root bridge, the BPDU is ignored and the port on Switch C transitions from a forwarding to a root-inconsistent state (shown by the green X icon). As a result, Switch A becomes the root bridge.

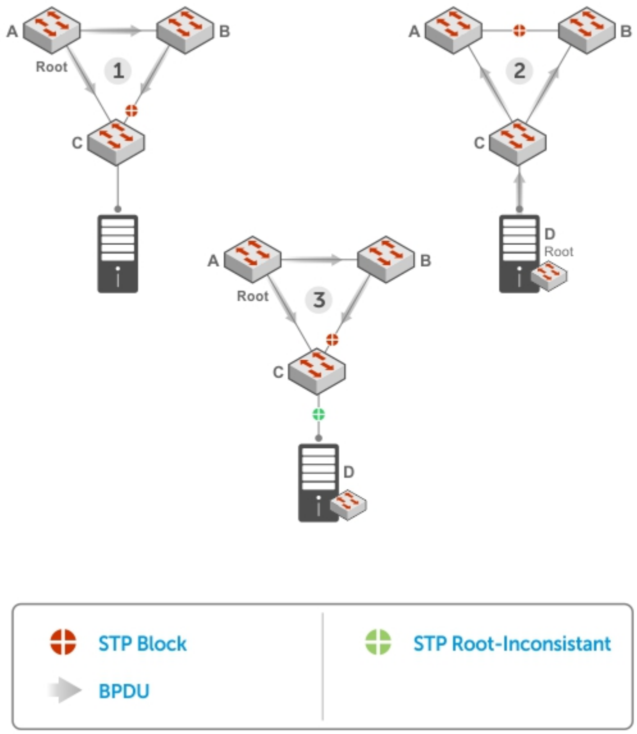


Figure 114. STP Root Guard Prevents Bridging Loops

Configuring Root Guard

Enable STP root guard on a per-port or per-port-channel basis.
Dell Networking OS Behavior: The following conditions apply to a port enabled with STP root guard:

- Root guard is supported on any STP-enabled port or port-channel interface except when used as a stacking port.
- Root guard is supported on a port in any Spanning Tree mode:
  - [Spanning Tree Protocol (STP)]
  - [Rapid Spanning Tree Protocol (RSTP)]
  - [Multiple Spanning Tree Protocol (MSTP)]
  - [Per-VLAN Spanning Tree Plus (PVST+)]
- When enabled on a port, root guard applies to all VLANs configured on the port.
- You cannot enable root guard and loop guard at the same time on an STP port. For example, if you configure root guard on a port on which loop guard is already configured, the following error message displays: • `% Error: LoopGuard is configured. Cannot configure RootGuard.`
- When used in an MSTP network, if root guard blocks a boundary port in the CIST, the port is also blocked in all other MST instances.

To enable the root guard on an STP-enabled port or port-channel interface in instance 0, use the following command.

- Enable root guard on a port or port-channel interface.

  INTERFACE mode or INTERFACE PORT-CHANNEL mode

  ```
  spanning-tree {0 | mstp | rstp | pvst} rootguard
  ```

  - `0`: enables root guard on an STP-enabled port assigned to instance 0.
  - `mstp`: enables root guard on an MSTP-enabled port.
  - `rstp`: enables root guard on an RSTP-enabled port.
  - `pvst`: enables root guard on a PVST-enabled port.

To disable STP root guard on a port or port-channel interface, use the `no spanning-tree 0 rootguard` command in an interface configuration mode.

To verify the STP root guard configuration on a port or port-channel interface, use the `show spanning-tree 0 guard [interface` *`interface`*`]` command in a global configuration mode.

Enabling SNMP Traps for Root Elections and Topology Changes

To enable SNMP traps individually or collectively, use the following commands.

- Enable SNMP traps for spanning tree state changes.

  `snmp-server enable traps stp`
- Enable SNMP traps for RSTP, MSTP, and PVST+ collectively.

  `snmp-server enable traps xstp`

Configuring Spanning Trees as Hitless

Configuring spanning trees as hitless is supported only on platform.
You can configure STP, RSTP, MSTP, and PVST+ to be hitless (configure all or none as hitless). When configured as hitless, critical protocol state information is synchronized between the RPMs so that RPM failover is seamless and no topology change is triggered.

To be hitless per spanning tree type or for all spanning tree types, use the following commands.

- Configure LACP to be hitless.

  CONFIGURATION mode

  ```
  redundancy protocol lacp
  ```
- Configure all spanning tree types to be hitless.

  CONFIGURATION mode

  ```
  redundancy protocol xstp
  ```

Example of Configuring all Spanning Tree Types to be Hitless

```
Dell(conf)#redundancy protocol xstp
Dell#show running-config redundancy
!
redundancy protocol xstp
Dell#
```

STP Loop Guard

STP loop guard is supported only on the platform.

The STP loop guard feature provides protection against Layer 2 forwarding loops (STP loops) caused by a hardware failure, such as a cable failure or an interface fault. When a cable or interface fails, a participating STP link may become unidirectional (STP requires links to be bidirectional) and an STP port does not receive BPDUs. When an STP blocking port does not receive BPDUs, it transitions to a Forwarding state. This condition can create a loop in the network.

For example, in the following example (STP topology 1, upper left), Switch A is the root switch and Switch B normally transmits BPDUs to Switch C. The link between Switch C and Switch B is in a Blocking state. However, if there is a unidirectional link failure (STP topology 1, lower left), Switch C does not receive BPDUs from Switch B. When the `max-age` timer expires, the STP port on Switch C becomes unblocked and transitions to Forwarding state. A loop is created as both Switch A and Switch C transmit traffic to Switch B.

As shown in the following illustration (STP topology 2, upper right), a loop can also be created if the forwarding port on Switch B becomes busy and does not forward BPDUs within the configured `forward-delay` time. As a result, the blocking port on Switch C transitions to a forwarding state, and both Switch A and Switch C transmit traffic to Switch B (STP topology 2, lower right).

As shown in STP topology 3 (bottom middle), after you enable loop guard on an STP port or port-channel on Switch C, if no BPDUs are received and the `max-age` timer expires, the port transitions from a blocked state to a Loop-Inconsistent state (instead of to a Forwarding state). Loop guard blocks the STP port so that no traffic is transmitted and no loop is created.

As soon as a BPDU is received on an STP port in a Loop-Inconsistent state, the port returns to a blocking state. If you disable STP loop guard on a port in a Loop-Inconsistent state, the port transitions to an STP blocking state and restarts the `max-age` timer.

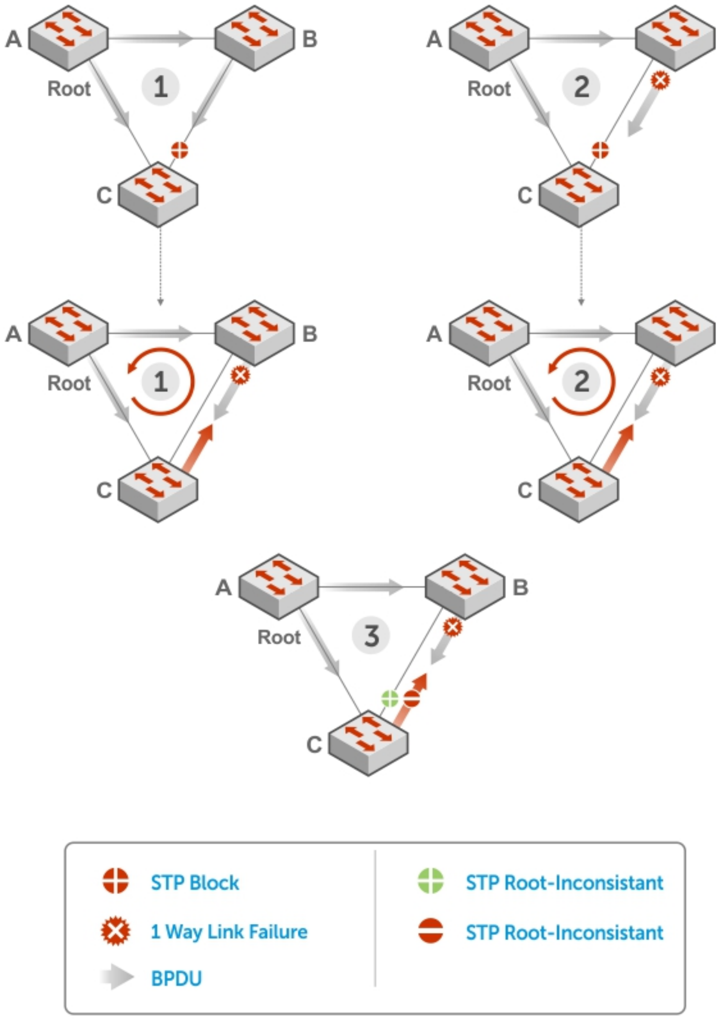

A
B
Root
C
1
2
3
STP Block
1 Way Link Failure
BPDU
STP Root-Inconsistant
STP Root-Inconsistant

Figure 115. STP Loop Guard Prevents Forwarding Loops

Configuring Loop Guard

Enable STP loop guard on a per-port or per-port channel basis.
Dell Networking OS Behavior: The following conditions apply to a port enabled with loop guard:

- Loop guard is supported on any STP-enabled port or port-channel interface.
- Loop guard is supported on a port or port-channel in any spanning tree mode:
  - Spanning Tree Protocol (STP)
  - Rapid Spanning Tree Protocol (RSTP)
  - Multiple Spanning Tree Protocol (MSTP)
  - Per-VLAN Spanning Tree Plus (PVST+)

- You cannot enable root guard and loop guard at the same time on an STP port. For example, if you configure loop guard on a port on which root guard is already configured, the following error message is displayed: `% Error: RootGuard is configured. Cannot configure LoopGuard`.
- Enabling Portfast BPDU guard and loop guard at the same time on a port results in a port that remains in a blocking state and prevents traffic from flowing through it. For example, when Portfast BPDU guard and loop guard are both configured:
  - If a BPDU is received from a remote device, BPDU guard places the port in an Err-Disabled Blocking state and no traffic is forwarded on the port.
  - If no BPDU is received from a remote device, loop guard places the port in a Loop-Inconsistent Blocking state and no traffic is forwarded on the port.
- When used in a PVST+ network, STP loop guard is performed per-port or per-port channel at a VLAN level. If no BPDUs are received on a VLAN interface, the port or port-channel transitions to a Loop-Inconsistent (Blocking) state only for this VLAN.

To enable a loop guard on an STP-enabled port or port-channel interface, use the following command.

- Enable loop guard on a port or port-channel interface.

  INTERFACE mode or INTERFACE PORT-CHANNEL mode

  ```
  spanning-tree {0 | mstp | rstp | pvst} loopguard
  ```

  - `0`: enables loop guard on an STP-enabled port assigned to instance 0.
  - `mstp`: enables loop guard on an MSTP-enabled port.
  - `rstp`: enables loop guard on an RSTP-enabled port.
  - `pvst`: enables loop guard on a PVST-enabled port.

To disable STP loop guard on a port or port-channel interface, use the `no spanning-tree 0 loopguard` command in an INTERFACE configuration mode.

To verify the STP loop guard configuration on a port or port-channel interface, use the `show spanning-tree 0 guard [interface *interface*]` command in a global configuration mode.

Displaying STP Guard Configuration

To display the STP guard configuration, use the following command.
The following example shows an STP network (instance 0) in which:

- Root guard is enabled on a port that is in a root-inconsistent state.
- Loop guard is enabled on a port that is in a listening state.
- BPDU guard is enabled on a port that is shut down (Error Disabled state) after receiving a BPDU.

- Verify the STP guard configured on port or port-channel interfaces.

  `show spanning-tree 0 guard [interface *interface*]`

Example of Viewing STP Guard Configuration

```
Dell#show spanning-tree 0 guard
Interface
Name    Instance Sts          Guard type
--------- -------- --------- ----------
Gi 0/1  0        INCON(Root)  Rootguard
Gi 0/2  0        LIS          Loopguard
Gi 0/3  0        EDS (Shut)   Bpduguard
```

47

System Time and Date

System time and date settings and the network time protocol (NTP) are supported on the S6000 platform.

You can set system times and dates and maintained through the NTP. They are also set through the Dell Networking Operating System (OS) command line interfaces (CLIs) and hardware settings.

In the release 9.4.(0.0), support for reaching an NTP server through different VRFs is included. You can configure a maximum of eight logging servers across different VRFs or the same VRF.

Network Time Protocol

The network time protocol (NTP) synchronizes timekeeping among a set of distributed time servers and clients.

The protocol also coordinates time distribution in a large, diverse network with various interfaces. In NTP, servers maintain the time and NTP clients synchronize with a time-serving host. NTP clients choose from among several NTP servers to determine which offers the best available source of time and the most reliable transmission of information.

NTP is a fault-tolerant protocol that automatically selects the best of several available time sources to synchronize to. You can combine multiple candidates to minimize the accumulated error. Temporarily or permanently insane time sources are detected and avoided.

Dell Networking recommends configuring NTP for the most accurate time. In Dell Networking OS, you can configure other time sources (the hardware clock and the software clock).

NTP is designed to produce three products: clock offset, roundtrip delay, and dispersion, all of which are relative to a selected reference clock.

- **Clock offset** — represents the amount to adjust the local clock to bring it into correspondence with the reference clock.
- **Roundtrip delay** — provides the capability to launch a message to arrive at the reference clock at a specified time.
- **Dispersion** — represents the maximum error of the local clock relative to the reference clock.

Because most host time servers synchronize via another peer time server, there are two components in each of these three products, those determined by the peer relative to the primary reference source of standard time and those measured by the host relative to the peer.

In order to facilitate error control and management of the subnet itself, each of these components is maintained separately in the protocol. They provide not only precision measurements of offset and delay, but also definitive maximum error bounds, so that the user interface can determine not only the time, but the quality of the time as well.

In what may be the most common client/server model, a client sends an NTP message to one or more servers and processes the replies as received. The server interchanges addresses and ports, overwrites certain fields in the message, recalculates the checksum and returns the message immediately. Information included in the NTP message allows the client to determine the server time regarding local time and adjust the local clock accordingly. In addition, the message includes information to calculate the expected timekeeping accuracy and reliability, as well as select the best from possibly several servers.

Following conventions established by the telephone industry [BEL86], the accuracy of each server is defined by a number called the stratum, with the topmost level (primary servers) assigned as one and each level downwards (secondary servers) in the hierarchy assigned as one greater than the preceding level.

Dell Networking OS synchronizes with a time-serving host to get the correct time. You can set Dell Networking OS to poll specific NTP time-serving hosts for the current time. From those time-serving hosts, the system chooses one NTP host with which to synchronize and serve as a client to the NTP host. As soon as a host-client relationship is established, the networking device propagates the time information throughout its local network.

Protocol Overview

The NTP messages to one or more servers and processes the replies as received. The server interchanges addresses and ports, fills in or overwrites certain fields in the message, recalculates the checksum, and returns it immediately.

Information included in the NTP message allows each client/server peer to determine the timekeeping characteristics of its other peers, including the expected accuracies of their clocks. Using this information, each peer is able to select the best time from possibly several other clocks, update the local clock, and estimate its accuracy.

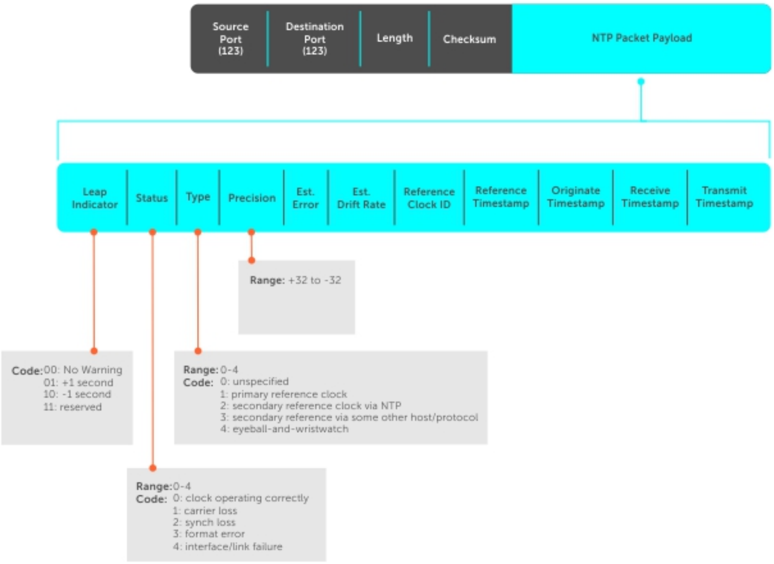

Source Port (123)
Destination Port (123)
Length
Checksum
NTP Packet Payload
Leap Indicator
Status
Type
Precision
Est. Error
Est. Drift Rate
Reference Clock ID
Reference Timestamp
Originate Timestamp
Receive Timestamp
Transmit Timestamp
Range: +32 to -32
Code: 00: No Warning
01: +1 second
10: -1 second
11: reserved
Range: 0-4
Code: 0: unspecified
1: primary reference clock
2: secondary reference clock via NTP
3: secondary reference via some other host/protocol
4: eyeball-and-wristwatch
Range: 0-4
Code: 0: clock operating correctly
1: carrier loss
2: synch loss
3: format error
4: interface/link failure

Figure 116. NTP Fields

Implementation Information

Dell Networking systems can only be an NTP client.

Configure the Network Time Protocol

Configuring NTP is a one-step process.

- Enabling NTP

Related Configuration Tasks

- Configuring NTP Broadcasts
- Setting the Hardware Clock with the Time Derived from NTP
- Disabling NTP on an Interface
- Configuring a Source IP Address for NTP Packets (optional)

Enabling NTP

NTP is disabled by default.
To enable NTP, specify an NTP server to which the Dell Networking system synchronizes. To specify multiple servers, enter the command multiple times. You may specify an unlimited number of servers at the expense of CPU resources.

- Specify the NTP server to which the Dell Networking system synchronizes.

  CONFIGURATION mode

  ```
  ntp server ip-address
  ```

Examples of Viewing System Clock

To display the system clock state with respect to NTP, use the `show ntp status` command from EXEC Privilege mode.

```
R6_E300(conf)#do show ntp status
Clock is synchronized, stratum 2, reference is 192.168.1.1
frequency is -369.623 ppm, stability is 53.319 ppm, precision is 4294967279
reference time is CD63BCC2.0CBBD000 (16:54:26.049 UTC Thu Mar 12 2009)
clock offset is 997.529984 msec, root delay is 0.00098 sec
root dispersion is 10.04271 sec, peer dispersion is 10032.715 msec
peer mode is client
```

To display the calculated NTP synchronization variables received from the server that the system uses to synchronize its clock, use the `show ntp associations` command from EXEC Privilege mode.

```
R6_E300(conf)#do show ntp associations
remote        ref clock st when poll reach delay offset disp
==========================================================
#192.168.1.1 .LOCL.     1 16 16 76         0.98  -2.470 879.23
* master (synced), # master (unsynced), + selected, - candidate
```

Setting the Hardware Clock with the Time Derived from NTP

To set the hardware clock, use the following command.

- Periodically update the system hardware clock with the time value derived from NTP.

CONFIGURATION mode

```
ntp update-calendar
```

Example of Updating the System Clock Relative to NTP

```
R5/R8(conf)#do show calendar
06:31:02 UTC Mon Mar 13 1989
R5/R8(conf)#ntp update-calendar 1
R5/R8(conf)#do show calendar
06:31:26 UTC Mon Mar 13 1989
R5/R8(conf)#do show calendar
12:24:11 UTC Thu Mar 12 2009
```

Configuring NTP Broadcasts

With Dell Networking OS, you can receive broadcasts of time information.
You can set interfaces within the system to receive NTP information through broadcast.

To configure an interface to receive NTP broadcasts, use the following commands.

- Set the interface to receive NTP packets.

  INTERFACE mode

  ```
  ntp broadcast client
  ```

Example of Configuring NTP Broadcasts

```
2w1d11h : NTP: Maximum Slew:-0.000470, Remainder = -0.496884
```

Disabling NTP on an Interface

By default, NTP is enabled on all active interfaces. If you disable NTP on an interface, Dell Networking OS drops any NTP packets sent to that interface.
To disable NTP on an interface, use the following command.

- Disable NTP on the interface.

  INTERFACE mode

  ```
  ntp disable
  ```

To view whether NTP is configured on the interface, use the `show config` command in INTERFACE mode. If `ntp disable` is not listed in the `show config` command output, NTP is enabled. (The `show config` command displays only non-default configuration information.)

Configuring a Source IP Address for NTP Packets

By default, the source address of NTP packets is the IP address of the interface used to reach the network.
You can configure one interface’s IP address include in all NTP packets.

To configure an IP address as the source address of NTP packets, use the following command.

- Configure a source IP address for NTP packets.

  CONFIGURATION mode

  ```
  ntp source interface
  ```

  Enter the following keywords and slot/port or number information:

- For a 1-Gigabit Ethernet interface, enter the keyword `GigabitEthernet` then the slot/port information.
- For a loopback interface, enter the keyword `loopback` then a number between 0 and 16383.
- For a port channel interface, enter the keyword `lag` then a number from 1 to 255 for TeraScale and ExaScale.
- For a SONET interface, enter the keyword `sonet` then the slot/port information.
- For a 10-Gigabit Ethernet interface, enter the keyword `TenGigabitEthernet` then the slot/port information.
- For a VLAN interface, enter the keyword `vlan` then a number from 1 to 4094.
- For a 40-Gigabit Ethernet interface, enter the keyword `fortyGigE` then the slot/port information.

To view the configuration, use the `show running-config ntp` command in EXEC privilege mode (refer to the example in Configuring NTP Authentication).

Configuring NTP Authentication

NTP authentication and the corresponding trusted key provide a reliable means of exchanging NTP packets with trusted time sources.
NTP authentication begins when the first NTP packet is created following the configuration of keys. NTP authentication in Dell Networking OS uses the message digest 5 (MD5) algorithm and the key is embedded in the synchronization packet that is sent to an NTP time source.

Dell Networking OS Behavior: Dell Networking OS uses an encryption algorithm to store the authentication key that is different from previous Dell Networking OS versions; Dell Networking OS uses data encryption standard (DES) encryption to store the key in the startup-config when you enter the `ntp authentication-key` command. Therefore, if your system boots with a startup-configuration from an Dell Networking OS version in which you have configured `ntp authentication-key`, the system cannot correctly decrypt the key and cannot authenticate the NTP packets. In this case, re-enter this command and save the running-config to the startup-config.

To configure NTP authentication, use the following commands.

1. Enable NTP authentication.

   CONFIGURATION mode

   ```
   ntp authenticate
   ```
2. Set an authentication key.

   CONFIGURATION mode

   ```
   ntp authentication-key number md5 key
   ```

   Configure the following parameters:
   - *`number`*: the range is from 1 to 4294967295. This *`number`* must be the same as the *`number`* in the `ntp trusted-key` command.
   - *`key`*: enter a text string. This text string is encrypted.
3. Define a trusted key.

   CONFIGURATION mode

   ```
   ntp trusted-key number
   ```

   Configure a number from 1 to 4294967295.

   The *`number`* must be the same as the *`number`* used in the `ntp authentication-key` command.

4. Configure an NTP server.

   CONFIGURATION mode

   ```
   ntp server ip-address [key keyid] [prefer] [version number]
   ```

   Configure the IP address of a server and the following optional parameters:

   - `key keyid`: configure a text string as the key exchanged between the NTP server and the client.
   - `prefer`: enter the keyword `prefer` to set this NTP server as the preferred server.
   - `version number`: enter a number as the NTP version. The range is from 1 to 3.

Examples of Configuring and Viewing an NTP Configuration

The following example shows configuring an NTP server.

```
R6_E300(conf)#1w6d23h : NTP: xmit packet to 192.168.1.1:
  leap 0, mode 3, version 3, stratum 2, ppoll 1024
  rtdel 0219 (8.193970), rtdsp AF928 (10973.266602), refid C0A80101
(192.168.1.1)
  ref CD7F4F63.6BE8F000 (14:51:15.421 UTC Thu Apr 2 2009)
  org CD7F4F63.68000000 (14:51:15.406 UTC Thu Apr 2 2009)
  rec CD7F4F63.6BE8F000 (14:51:15.421 UTC Thu Apr 2 2009)
  xmt CD7F5368.D0535000 (15:8:24.813 UTC Thu Apr 2 2009)
1w6d23h : NTP: rcv packet from 192.168.1.1
  leap 0, mode 4, version 3, stratum 1, ppoll 1024
  rtdel 0000 (0.000000), rtdsp AF587 (10959.090820), refid 4C4F434C
(76.79.67.76)
  ref CD7E14FD.43F7CED9 (16:29:49.265 UTC Wed Apr 1 2009)
  org CD7F5368.D0535000 (15:8:24.813 UTC Thu Apr 2 2009)
  rec CD7F5368.D0000000 (15:8:24.812 UTC Thu Apr 2 2009)
  xmt CD7F5368.D0000000 (15:8:24.812 UTC Thu Apr 2 2009)
  inp CD7F5368.D1974000 (15:8:24.818 UTC Thu Apr 2 2009)

rtdel-root delay
rtdsp - round trip dispersion
refid - reference id
org -
rec - (last?) receive timestamp
xmt - transmit timestamp

mode - 3 client, 4 server
stratum - 1 primary reference clock, 2 secondary reference clock (via NTP)
version - NTP version 3
leap -
```

NOTE:

- **Leap Indicator** (`sys.leap`, `peer.leap`, `pkt.leap`) — This is a two-bit code warning of an impending leap second to be inserted in the NTP time scale. The bits are set before 23:59 on the day of insertion and reset after 00:00 on the following day. This causes the number of seconds (rollover interval) in the day of insertion to be increased or decreased by one. In the case of primary servers, the bits are set by operator intervention, while in the case of secondary servers, the bits are set by the protocol. The two bits, bit 0, and bit 1, respectively, are coded as follows:
- **Poll Interval** — integer indicating the minimum interval between transmitted messages, in seconds as a power of two. For instance, a value of six indicates a minimum interval of 64 seconds.
- **Precision** — integer indicating the precision of the various clocks, in seconds to the nearest power of two. The value must be rounded to the next larger power of two; for instance, a 50 Hz (20 ms) or 60 Hz (16.67ms) power-frequency clock is assigned the value -5 (31.25 ms), while a 1000 Hz (1 ms) crystal-controlled clock is assigned the value -9 (1.95 ms).
- **Root Delay** (`sys.rootdelay`, `peer.rootdelay`, `pkt.rootdelay`) — a signed fixed-point number indicating the total round-trip delay to the primary reference source at the root of the synchronization subnet, in seconds. This variable can take on both positive and negative values, depending on clock precision and skew.
- **Root Dispersion** (`sys.rootdispersion`, `peer.rootdispersion`, `pkt.rootdispersion`) — a signed fixed-point number indicating the maximum error relative to the primary reference source at the root of the synchronization subnet, in seconds. Only positive values greater than zero are possible.
- **Reference Clock Identifier** (`sys.refid`, `peer.refid`, `pkt.refid`) — This is a 32-bit code identifying the particular reference clock. In the case of stratum 0 (unspecified) or stratum 1 (primary reference source), this is a four-octet, left-justified, zero-padded ASCII string, for example: in the case of stratum 2 and greater (secondary reference) this is the four-octet internet address of the peer selected for synchronization.
- **Reference Timestamp** (`sys.reftime`, `peer.reftime`, `pkt.reftime`) — This is the local time, in timestamp format, when the local clock was last updated. If the local clock has never been synchronized, the value is zero.
- `Originate Timestamp`: The departure time on the server of its last NTP message. If the server becomes unreachable, the value is set to zero.
- **Receive Timestamp** — the arrival time on the client of the last NTP message from the server. If the server becomes unreachable, the value is set to zero.
- **Transmit Timestamp** — the departure time on the server of the current NTP message from the sender.
- **Filter dispersion** — the error in calculating the minimum delay from a set of sample data from a peer.

To view the NTP configuration, use the `show running-config ntp` command in EXEC privilege mode. The following example shows an encrypted authentication key (in bold). All keys are encrypted.

```
Dell#show running ntp
!
ntp authenticate
ntp authentication-key 345 md5 5A60910F3D211F02
ntp server 11.1.1.1 version 3
ntp trusted-key 345
Dell#
```

Dell Networking OS Time and Date

You can set the time and date using the Dell Networking OS CLI.

Configuration Task List

The following is a configuration task list for configuring the time and date settings.

- Setting the Time and Date for the Switch Hardware Clock
- Setting the Time and Date for the Switch Software Clock
- Setting the Timezone
- Setting Daylight Saving Time Once
- Setting Recurring Daylight Saving Time

Setting the Time and Date for the Switch Hardware Clock

To set the time and date for the switch hardware clock, use the following command.

- Set the hardware clock to the current time and date.

  EXEC Privilege mode

  ```
  calendar set time month day year
  ```

  - `time`: enter the time in hours:minutes:seconds. For the hour variable, use the 24-hour format; for example, 17:15:00 is 5:15 pm.
  - `month`: enter the name of one of the 12 months in English. You can enter the name of a day to change the order of the display to time day month year.
  - `day`: enter the number of the day. The range is from 1 to 31. You can enter the name of a month to change the order of the display to time day month year.
  - `year`: enter a four-digit number as the year. The range is from 1993 to 2035.

Example of the `calendar set` Command

```
Dell#calendar set 08:55:00 september 18 2009
Dell#
```

Setting the Time and Date for the Switch Software Clock

You can change the order of the `month` and `day` parameters to enter the time and date as *time day month year*. You cannot delete the software clock.
The software clock runs only when the software is up. The clock restarts, based on the hardware clock, when the switch reboots.

To set the software clock, use the following command.

- Set the system software clock to the current time and date.

  EXEC Privilege mode

  ```
  clock set time month day year
  ```

  - `time`: enter the time in hours:minutes:seconds. For the hour variable, use the 24-hour format; for example, 17:15:00 is 5:15 pm.

  - *`month`*: enter the name of one of the 12 months in English. You can enter the name of a day to change the order of the display to *time day month year*.
  - *`day`*: enter the number of the day. The range is from 1 to 31. You can enter the name of a month to change the order of the display to *time day month year*.
  - *`year`*: enter a four-digit number as the year. The range is from 1993 to 2035.

Example of the `clock set` Command

```
Dell#clock set 16:20:00 19 september 2009
Dell#
```

Setting the Timezone

Universal time coordinated (UTC) is the time standard based on the International Atomic Time standard, commonly known as Greenwich Mean time.
When determining system time, include the differentiator between UTC and your local timezone. For example, San Jose, CA is the Pacific Timezone with a UTC offset of -8.

To set the clock timezone, use the following command.

- Set the clock to the appropriate timezone.

  CONFIGURATION mode

  ```
  clock timezone timezone-name offset
  ```

  - *`timezone-name`*: enter the name of the timezone. Do not use spaces.
  - *`offset`*: enter one of the following:
    - a number from 1 to 23 as the number of hours in addition to UTC for the timezone.
    - a minus sign (-) then a number from 1 to 23 as the number of hours.

Example of the `clock timezone` Command

```
Dell#conf
Dell(conf)#clock timezone Pacific -8
Dell(conf)#01:40:19: %RPM0-P:CP %CLOCK-6-TIME CHANGE: Timezone
configuration changed from "UTC 0 hrs 0 mins" to "Pacific -8 hrs 0
mins"
Dell#
```

Set Daylight Saving Time

Dell Networking OS supports setting the system to daylight saving time once or on a recurring basis every year.

Setting Daylight Saving Time Once

Set a date (and time zone) on which to convert the switch to daylight saving time on a one-time basis.
To set the clock for daylight savings time once, use the following command.

- Set the clock to the appropriate timezone and daylight saving time.

  CONFIGURATION mode

  ```
  clock summer-time time-zone date start-month start-day start-year start-time
  end-month end-day end-year end-time [offset]
  ```

- *`time-zone`*: enter the three-letter name for the time zone. This name displays in the show clock output.
- *`start-month`*: enter the name of one of the 12 months in English. You can enter the name of a day to change the order of the display to *time day month year*.
- *`start-day`*: enter the number of the day. The range is from 1 to 31. You can enter the name of a month to change the order of the display to *time day month year*.
- *`start-year`*: enter a four-digit number as the year. The range is from 1993 to 2035.
- *`start-time`*: enter the time in hours:minutes. For the hour variable, use the 24-hour format; example, 17:15 is 5:15 pm.
- *`end-month`*: enter the name of one of the 12 months in English. You can enter the name of a day to change the order of the display to *time day month year*.
- *`end-day`*: enter the number of the day. The range is from 1 to 31. You can enter the name of a month to change the order of the display to *time day month year*.
- *`end-year`*: enter a four-digit number as the year. The range is from 1993 to 2035.
- *`end-time`*: enter the time in hours:minutes. For the hour variable, use the 24-hour format; example, 17:15 is 5:15 pm.
- *`offset`*: (OPTIONAL) enter the number of minutes to add during the summer-time period. The range is from 1 to1440. The default is **60 minutes**.

Example of the `clock summer-time` Command

```
Dell(conf)#clock summer-time pacific date Mar 14 2009 00:00 Nov 7 2009 00:00
Dell(conf)#02:02:13: %RPM0-P:CP %CLOCK-6-TIME CHANGE: Summertime configuration
changed from
"none" to "Summer time starts 00:00:00 Pacific Sat Mar 14 2009;Summer time ends
00:00:00 pacific
Sat Nov 7 2009"
```

Setting Recurring Daylight Saving Time

Set a date (and time zone) on which to convert the switch to daylight saving time on a specific day every year.
If you have already set daylight saving for a one-time setting, you can set that date and time as the recurring setting with the `clock summer-time time-zone recurring` command.

To set a recurring daylight saving time, use the following command.

- Set the clock to the appropriate timezone and adjust to daylight saving time every year.

  CONFIGURATION mode

  ```
  clock summer-time time-zone recurring start-week start-day start-month start-
  time end-week end-day end-month end-time [offset]
  ```

  - *`time-zone`*: Enter the three-letter name for the time zone. This name displays in the show clock output.
  - *`start-week`*: (OPTIONAL) Enter one of the following as the week that daylight saving begins and then enter values for *`start-day`* through *`end-time`*:
    - *`week-number`*: Enter a number from 1 to 4 as the number of the week in the month to start daylight saving time.
    - `first`: Enter the keyword `first` to start daylight saving time in the first week of the month.
    - `last`: Enter the keyword `last` to start daylight saving time in the last week of the month.
  - *`start-month`*: Enter the name of one of the 12 months in English. You can enter the name of a day to change the order of the display to *time day month year*.

- `start-day`: Enter the number of the day. The range is from 1 to 31. You can enter the name of a month to change the order of the display to *time day month year*.
- `start-year`: Enter a four-digit number as the year. The range is from 1993 to 2035.
- `start-time`: Enter the time in hours:minutes. For the hour variable, use the 24-hour format; example, 17:15 is 5:15 pm.
- `end-week`: If you entered a start-week, enter the one of the following as the week that daylight saving ends:
  - `week-number`: Enter a number from 1 to 4 as the number of the week in the month to start daylight saving time.
  - `first`: Enter the keyword `first` to start daylight saving time in the first week of the month.
  - `last`: Enter the keyword `last` to start daylight saving time in the last week of the month.
- `end-month`: Enter the name of one of the 12 months in English. You can enter the name of a day to change the order of the display to *time day month year*.
- `end-day`: Enter the number of the day. The range is from 1 to 31. You can enter the name of a month to change the order of the display to *time day month year*.
- `end-year`: Enter a four-digit number as the year. The range is from 1993 to 2035.
- `end-time`: Enter the time in hours:minutes. For the hour variable, use the 24-hour format; example, 17:15 is 5:15 pm.
- `offset`: (OPTIONAL) Enter the number of minutes to add during the summer-time period. The range is from 1 to1440. The default is **60 minutes**.

Examples of the `clock summer-time recurring` Command

The following example shows the `clock summer-time recurring` command.

```
Dell(conf)#clock summer-time pacific recurring Mar 14 2009 00:00 Nov 7 2009
00:00 ?
Dell(conf)#02:02:13: %RPM0-P:CP %CLOCK-6-TIME CHANGE: Summertime configuration
changed from
"none" to "Summer time starts 00:00:00 Pacific Sat Mar 14 2009;Summer time ends
00:00:00 pacific
Sat Nov 7 2009"
```

NOTE: If you enter <CR> after entering the `recurring` command parameter, and you have already set a one-time daylight saving time/date, the system uses that time and date as the recurring setting.

The following example shows the clock summer-time recurring parameters.

```
Dell(conf)#clock summer-time pacific recurring ?
<1-4>     Week number to start
first     Week number to start
last      Week number to start
<cr>
Dell(conf)#clock summer-time pacific recurring
Dell(conf)#02:10:57: %RPM0-P:CP %CLOCK-6-TIME CHANGE: Summertime configuration
changed from
"Summer time starts 00:00:00 Pacific Sat Mar 14 2009 ; Summer time ends
00:00:00 pacific Sat Nov
7 2009" to "Summer time starts 02:00:00 Pacific Sun Mar 8 2009;Summer time ends
02:00:00 pacific
Sun Nov 1 2009"
```

48

Tunneling

Tunnel interfaces create a logical tunnel for IPv4 or IPv6 traffic. Tunneling supports RFC 2003, RFC 2473, and 4213.

DSCP, hop-limits, flow label values, OSPFv2, and OSPFv3 are also supported. ICMP error relay, PATH MTU transmission, and fragmented packets are not supported.

Configuring a Tunnel

You can configure a tunnel in IPv6 mode, IPv6IP mode, and IPIP mode.
You can configure a tunnel in IPv6 mode, IPv6IP mode, and IPIP mode.

- If the tunnel mode is IPIP or IPv6IP, the tunnel source address and the tunnel destination address must be an IPv4 address.
- If the tunnel mode is IPv6, the tunnel source address and the tunnel destination address must be an IPv6 address.
- If the tunnel mode is IPv6 or IPIP, you can use either an IPv6 address or an IPv4 address for the logical address of the tunnel, but in IPv6IP mode, the logical address must be an IPv6 address.

The following sample configuration shows a tunnel configured in IPv6 mode (carries IPv6 and IPv4 traffic).

```
Dell(conf)#interface tunnel 1
Dell(conf-if-tu-1)#tunnel source 30.1.1.1
Dell(conf-if-tu-1)#tunnel destination 50.1.1.1
Dell(conf-if-tu-1)#tunnel mode ipip
Dell(conf-if-tu-1)#ip address 1.1.1.1/24
Dell(conf-if-tu-1)#ipv6 address 1::1/64
Dell(conf-if-tu-1)#no shutdown
Dell(conf-if-tu-1)#show config
!
interface Tunnel 1
ip address 1.1.1.1/24
ipv6 address 1::1/64
tunnel destination 50.1.1.1
tunnel source 30.1.1.1
tunnel mode ipip
no shutdown
```

The following sample configuration shows a tunnel configured in IPV6IP mode (IPv4 tunnel carries IPv6 traffic only):

```
Dell(conf)#interface tunnel 2
Dell(conf-if-tu-2)#tunnel source 60.1.1.1
Dell(conf-if-tu-2)#tunnel destination 90.1.1.1
Dell(conf-if-tu-2)#tunnel mode ipv6ip
Dell(conf-if-tu-2)#ipv6 address 2::1/64
Dell(conf-if-tu-2)#no shutdown
Dell(conf-if-tu-2)#show config
!
interface Tunnel 2
no ip address
```

```
ipv6 address 2::1/64
tunnel destination 90.1.1.1
tunnel source 60.1.1.1
tunnel mode ipv6ip
no shutdown
```

The following sample configuration shows a tunnel configured in IPIP mode (IPv4 tunnel carries IPv4 and IPv6 traffic):

```
Dell(conf)#interface tunnel 3
Dell(conf-if-tu-3)#tunnel source 5::5
Dell(conf-if-tu-3)#tunnel destination 8::9
Dell(conf-if-tu-3)#tunnel mode ipv6
Dell(conf-if-tu-3)#ip address 3.1.1.1/24
Dell(conf-if-tu-3)#ipv6 address 3::1/64
Dell(conf-if-tu-3)#no shutdown
Dell(conf-if-tu-3)#show config
!
interface Tunnel 3
ip address 3.1.1.1/24
ipv6 address 3::1/64
tunnel destination 8::9
tunnel source 5::5
tunnel mode ipv6
no shutdown
```

Configuring Tunnel Keepalive Settings

You can configure a tunnel keepalive target, keepalive interval, and attempts.

NOTE: By default the tunnel keepalive is disabled.

The following sample configuration shows how to use tunnel keepalive command.

```
Dell(conf-if-te-0/12)#show config
!
interface TenGigabitEthernet 0/12
ip address 40.1.1.1/24
ipv6 address 500:10::1/64
no shutdown
Dell(conf-if-te-0/12)#
Dell(conf)#interface tunnel 1
Dell(conf-if-tu-1)#ipv6 address 1abd::1/64
Dell(conf-if-tu-1)#ip address 1.1.1.1/24
Dell(conf-if-tu-1)#tunnel source 40.1.1.1
Dell(conf-if-tu-1)#tunnel destination 40.1.1.2
Dell(conf-if-tu-1)#tunnel mode ipip
Dell(conf-if-tu-1)#no shutdown
Dell(conf-if-tu-1)#tunnel keepalive 1.1.1.2 attempts 4 interval 6
Dell(conf-if-tu-1)#show config
!
interface Tunnel 1
ip address 1.1.1.1/24
ipv6 address 1abd::1/64
tunnel destination 40.1.1.2
tunnel source 40.1.1.1
tunnel keepalive 1.1.1.2 attempts 4 interval 6
tunnel mode ipip
no shutdown
```

Configuring a Tunnel Interface

You can configure the tunnel interface using the `ip unnumbered` and `ipv6 unnumbered` commands.

To configure the tunnel interface to operate without a unique explicit ip or ipv6 address, select the interface from which the tunnel will borrow its address.

The following sample configuration shows how to use the tunnel interface configuration commands.

```
Dell(conf-if-te-0/0)#show config
 !
interface TenGigabitEthernet 0/0
ip address 20.1.1.1/24
ipv6 address 20:1::1/64
no shutdown
Dell(conf)#interface tunnel 1
Dell(conf-if-tu-1)#ip unnumbered tengigabitethernet 0/0
Dell(conf-if-tu-1)#ipv6 unnumbered tengigabitethernet 0/0
Dell(conf-if-tu-1)#tunnel source 40.1.1.1
Dell(conf-if-tu-1)#tunnel mode ipip decapsulate-any
Dell(conf-if-tu-1)#no shutdown
Dell(conf-if-tu-1)#show config
!
interface Tunnel 1
ip unnumbered TenGigabitEthernet 0/0
ipv6 unnumbered TenGigabitEthernet 0/0
tunnel source 40.1.1.1
tunnel mode ipip decapsulate-any
no shutdown
Dell(conf-if-tu-1)#
```

Configuring Tunnel allow-remote Decapsulation

You can configure an IPv4 or IPV6 address or prefix whose tunneled packet will be accepted for decapsulation.

- If no allow-remote entries are configured, then tunneled packets from any remote peer address will be accepted.
- Upto eight allow-remote entries can be configured on any particular multipoint receive-only tunnel.

The following sample configuration shows how to configure a tunnel allow-remote address.

```
Dell(conf)#interface tunnel 1
Dell(conf-if-tu-1)#ipv6 address 1abd::1/64
Dell(conf-if-tu-1)#ip address 1.1.1.1/24
Dell(conf-if-tu-1)#tunnel source 40.1.1.1
Dell(conf-if-tu-1)#tunnel mode ipip decapsulate-any
Dell(conf-if-tu-1)#tunnel allow-remote 40.1.1.2
Dell(conf-if-tu-1)#no shutdown
Dell(conf-if-tu-1)#show config
!
interface Tunnel 1
ip address 1.1.1.1/24
ipv6 address 1abd::1/64
tunnel source 40.1.1.1
tunnel allow-remote 40.1.1.2
tunnel mode ipip decapsulate-any
no shutdown
```

Configuring Tunnel source anylocal Decapsulation

The `tunnel source anylocal` command allows a multipoint receive-only tunnel to decapsulate tunnel packets addressed to any IPv4 or IPv6 (depending on the tunnel mode) address configured on the switch that is operationally UP.

The `source anylocal` parameters can be used for packet decapsulation instead of the ip address or interface (`tunnel allow-remote` command), but only on multipoint receive-only mode tunnels.

The following sample configuration shows how to use the `tunnel source anylocal` command.

```
Dell(conf)#interface tunnel 1
Dell(conf-if-tu-1)#ipv6 address 1abd::1/64
Dell(conf-if-tu-1)#ip address 1.1.1.1/24
Dell(conf-if-tu-1)#tunnel source anylocal
Dell(conf-if-tu-1)#tunnel mode ipip decapsulate-any
Dell(conf-if-tu-1)#tunnel allow-remote 40.1.1.2
Dell(conf-if-tu-1)#no shutdown
Dell(conf-if-tu-1)#show config
!
interface Tunnel 1
ip address 1.1.1.1/24
ipv6 address 1abd::1/64
tunnel source anylocal
tunnel allow-remote 40.1.1.2
tunnel mode ipip decapsulate-any
no shutdown
```

Guidelines for Configuring Multipoint Receive-Only Tunnels

- Maximum number of allowed remote end-points that can be configured for a single multipoint receive-only tunnel is eight. Maximum number of allowed remote end-points that can be configured for all multipoint receive-only tunnels depends on the hardware table size to setup termination (it is 512 entries in S4810 and S4820T platforms) and the count is tracked across all of the tunnel remote end-points configured in the system.
- With multipoint receive-only tunnels, hardware should be handling multiple nested encapsulated packets and this need to work with the IP MTU set on the physical interface.
- Control plane packets received on the multipoint-receive-only tunnel, are packets destined to the local ip address and routed to the CPU after decapsulation, would follow the regular processing in the TCP/IP stack. A response to these packets is only possible if the route back to the sender does not fall on a receive-only tunnel.
- Multipath over more than one VLAN interfaces will not be working for packets routed over the tunnel interface. This is inherited from the current implementation and this is not applicable for the receive-only tunnels functionality discussed above.
- IP tunnel interfaces are supported over ECMP paths to the next hop. ECMP paths over IP tunnel interfaces are supported. ARP and neighbor resolution for the IP tunnel next hop are supported.

Multipoint Receive-Only Type and IP Unnumbered Interfaces for Tunnels

Multipoint receive-only type IP Tunnel is now supported in S6000 platform. This is a new type of tunnel that is expected to only decapsulate packets from remote end points but never forwards packets on the tunnel. Additional level of protection on the receive-only type IP tunnels is available by allowing only a given prefix/range of remote peers. IP unnumbered interface address configuration on the IP Tunnels is now supported.

Tunnel interface's operational status will always remain UP once tunnel is configured to work in multipoint-receive-only mode. Allowed remote addresses configured for multipoint-receive-only type tunnel are setup only for decapsulation and hence they are not marked for neighbor resolution like the regular tunnel's destination address. Connected route for the tunnel interface's IP subnet will not be pointing towards the tunnel and it will be pointing towards CPU for the receive-only tunnels. Tunnel interface can function as IP unnumbered interface with no IP address assigned on it

49

Upgrade Procedures

To find the upgrade procedures, go to the *Dell Networking OS Release Notes* for your system type to see all the requirements needed to upgrade to the desired Dell Networking OS version. To upgrade your system type, follow the procedures in the *Dell Networking OS Release Notes*.

Get Help with Upgrades

Direct any questions or concerns about the Dell Networking OS upgrade procedures to the Dell Technical Support Center. You can reach Technical Support:

- On the web: http://www.dell.com/support
- By email: Dell-Force10_Technical_Support@Dell.com
- By phone: US and Canada: 866.965.5800, International: 408.965.5800.

50

Virtual LANs (VLANs)

Virtual LANs (VLANs) are supported on the S6000 platform.

VLANs are a logical broadcast domain or logical grouping of interfaces in a local area network (LAN) in which all data received is kept locally and broadcast to all members of the group. When in Layer 2 mode, VLANs move traffic at wire speed and can span multiple devices. The Dell Networking Operating System (OS) supports up to 4093 port-based VLANs and one default VLAN, as specified in IEEE 802.1Q.

VLANs benefits include:

- Improved security because you can isolate groups of users into different VLANs
- Ability to create one VLAN across multiple devices

For more information about VLANs, refer to the *IEEE Standard 802.1Q Virtual Bridged Local Area Networks*. In this guide, also refer to:

- Bulk Configuration in the Interfaces chapter.
- VLAN Stacking in the Service Provider Bridging chapter.

For a complete listing of all commands related to Dell Networking OS VLANs, refer to these *Dell Networking OS Command Reference Guide* chapters:

- Interfaces
- 802.1X
- GARP VLAN Registration Protocol (GVRP)
- Service Provider Bridging
- Per-VLAN Spanning Tree Plus (PVST+)

The following table lists the defaults for VLANs in Dell Networking OS.

Feature	Default
Spanning Tree group ID	All VLANs are part of Spanning Tree group 0.
Mode	Layer 2 (no IP address is assigned).
Default VLAN ID	VLAN 1

Default VLAN

When you configure interfaces for Layer 2 mode, they are automatically placed in the Default VLAN as untagged interfaces. Only untagged interfaces can belong to the Default VLAN.

The following example displays the outcome of placing an interface in Layer 2 mode. To configure an interface for Layer 2 mode, use the `switchport` command. As shown in bold, the `switchport` command places the interface in Layer 2 mode and the `show vlan` command in EXEC privilege mode indicates that the interface is now part of the Default VLAN (VLAN 1).

By default, VLAN 1 is the Default VLAN. To change that designation, use the `default vlan-id` command in CONFIGURATION mode. You cannot delete the Default VLAN.

NOTE: You cannot assign an IP address to the Default VLAN. To assign an IP address to a VLAN that is currently the Default VLAN, create another VLAN and assign it to be the Default VLAN. For more information about assigning IP addresses, refer to Assigning an IP Address to a VLAN.

- Untagged interfaces must be part of a VLAN. To remove an untagged interface from the Default VLAN, create another VLAN and place the interface into that VLAN. Alternatively, use the `no switchport` command, and Dell Networking OS removes the interface from the Default VLAN.
- A tagged interface requires an additional step to remove it from Layer 2 mode. Because tagged interfaces can belong to multiple VLANs, remove the tagged interface from all VLANs using the `no tagged interface` command. Only after the interface is untagged and a member of the Default VLAN can you use the `no switchport` command to remove the interface from Layer 2 mode. For more information, refer to VLANs and Port Tagging.

Example of Configuring an Interface for Layer 2 Belonging to the Default VLAN

```
Dell(conf)#int gi 3/2
Dell(conf-if)#no shut
Dell(conf-if)#switchport
Dell(conf-if)#show config
!
interface GigabitEthernet 3/2
  no ip address
  switchport
  no shutdown
Dell(conf-if)#end
Dell#show vlan

Codes: * - Default VLAN, G - GVRP VLANs
  NUM   Status   Q Ports
* 1     Active   U Gi 3/2
  2     Active   T Po1(So 0/0-1)
                T Gi 3/0
Dell#
```

Port-Based VLANs

Port-based VLANs are a broadcast domain defined by different ports or interfaces. In Dell Networking OS, a port-based VLAN can contain interfaces from different line cards within the chassis. Dell Networking OS supports 4094 port-based VLANs.

Port-based VLANs offer increased security for traffic, conserve bandwidth, and allow switch segmentation. Interfaces in different VLANs do not communicate with each other, adding some security to the traffic on those interfaces. Different VLANs can communicate between each other by means of IP routing. Because traffic is only broadcast or flooded to the interfaces within a VLAN, the VLAN conserves bandwidth. Finally, you can have multiple VLANs configured on one switch, thus segmenting the device.

Interfaces within a port-based VLAN must be in Layer 2 mode and can be tagged or untagged in the VLAN ID.

VLANs and Port Tagging

To add an interface to a VLAN, the interface must be in Layer 2 mode. After you place an interface in Layer 2 mode, the interface is automatically placed in the Default VLAN.

Dell Networking OS supports IEEE 802.1Q tagging at the interface level to filter traffic. When you enable tagging, a tag header is added to the frame after the destination and source MAC addresses. That information is preserved as the frame moves through the network. The following example shows the structure of a frame with a tag header. The VLAN ID is inserted in the tag header.

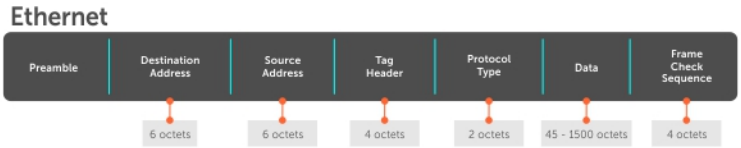


Figure 117. Tagged Frame Format

The tag header contains some key information that Dell Networking OS uses:

- The VLAN protocol identifier identifies the frame as tagged according to the IEEE 802.1Q specifications (2 bytes).
- Tag control information (TCI) includes the VLAN ID (2 bytes total). The VLAN ID can have 4,096 values, but two are reserved.

NOTE: The insertion of the tag header into the Ethernet frame increases the size of the frame to more than the 1,518 bytes as specified in the IEEE 802.3 standard. Some devices that are not compliant with IEEE 802.3 may not support the larger frame size.

Information contained in the tag header allows the system to prioritize traffic and to forward information to ports associated with a specific VLAN ID. Tagged interfaces can belong to multiple VLANs, while untagged interfaces can belong only to one VLAN.

Configuration Task List

This section contains the following VLAN configuration tasks.

- Creating a Port-Based VLAN (mandatory)
- Assigning Interfaces to a VLAN (optional)
- Assigning an IP Address to a VLAN (optional)
- Enabling Null VLAN as the Default VLAN

Creating a Port-Based VLAN

To configure a port-based VLAN, create the VLAN and then add physical interfaces or port channel (LAG) interfaces to the VLAN.

NOTE: The Default VLAN (VLAN 1) is part of the system startup configuration and does not require configuration.

A VLAN is active only if the VLAN contains interfaces and those interfaces are operationally up. As shown in the following example, VLAN 1 is inactive because it does not contain any interfaces. The other VLANs contain enabled interfaces and are active.

NOTE: In a VLAN, the `shutdown` command stops Layer 3 (routed) traffic only. Layer 2 traffic continues to pass through the VLAN. If the VLAN is not a routed VLAN (that is, configured with an IP address), the `shutdown` command has no affect on VLAN traffic.

When you delete a VLAN (using the `no interface vlan` *`vlan-id`* command), any interfaces assigned to that VLAN are assigned to the Default VLAN as untagged interfaces.

To create a port-based VLAN, use the following command.

- Configure a port-based VLAN (if the VLAN-ID is different from the Default VLAN ID) and enter INTERFACE VLAN mode.

  CONFIGURATION mode

  `interface vlan *vlan-id*`

  To activate the VLAN, after you create a VLAN, assign interfaces in Layer 2 mode to the VLAN.

Example of Verifying a Port-Based VLAN

To view the configured VLANs, use the `show vlan` command in EXEC Privilege mode.

```
Dell#show vlan

Codes: * - Default VLAN, G - GVRP VLANs

  NUM  Status   Q  Ports
* 1    Inactive U  So 9/4-11
  2    Active   U  Gi 0/1,18
  3    Active   U  Gi 0/2,19
  4    Active   T  Gi 0/3,20
  5    Active   U  Po 1
  6    Active   U  Gi 0/12
                U  So 9/0
Dell#
```

Assigning Interfaces to a VLAN

You can only assign interfaces in Layer 2 mode to a VLAN using the tagged and untagged commands. To place an interface in Layer 2 mode, use the `switchport` command.
You can further designate these Layer 2 interfaces as tagged or untagged. For more information, refer to the Interfaces chapter and Configuring Layer 2 (Data Link) Mode. When you place an interface in Layer 2 mode by the `switchport` command, the interface is automatically designated untagged and placed in the Default VLAN.

To view which interfaces are tagged or untagged and to which VLAN they belong, use the `show vlan` command. The following example shows that six VLANs are configured, and two interfaces are assigned to VLAN 2. The Q column in the `show vlan` command example notes whether the interface is tagged (T) or untagged (U). For more information about this command, refer to the Layer 2 chapter of the *Dell Networking OS Command Reference Guide*.

To tag frames leaving an interface in Layer 2 mode, assign that interface to a port-based VLAN to tag it with that VLAN ID. To tag interfaces, use the following commands.

1. Access INTERFACE VLAN mode of the VLAN to which you want to assign the interface.

   CONFIGURATION mode

   `interface vlan *vlan-id*`
2. Enable an interface to include the IEEE 802.1Q tag header.

   INTERFACE mode

   `tagged *interface*`

Add an Interface to Another VLAN

To view just the interfaces that are in Layer 2 mode, use the `show interfaces switchport` command in EXEC Privilege mode or EXEC mode.

The following example shows the steps to add a tagged interface (in this case, port channel 1) to VLAN 4. To view the interface's status. Interface (po 1) is tagged and in VLAN 2 and 3, use the `show vlan` command. In a port-based VLAN, use the tagged command to add the interface to another VLAN. The `show vlan` command output displays the interface's (po 1) changed status.

Except for hybrid ports, only a tagged interface can be a member of multiple VLANs. You can assign hybrid ports to two VLANs if the port is untagged in one VLAN and tagged in all others.

```
Dell#show vlan

Codes: * - Default VLAN, G - GVRP VLANs

  NUM Status   Q  Ports
* 1   Inactive
  2   Active   T  Po1(So 0/0-1)
               T  Gi 3/0
  3   Active   T  Po1(So 0/0-1)
               T  Gi 3/1

Dell#config
Dell(conf)#int vlan 4
Dell(conf-if-vlan)#tagged po 1
Dell(conf-if-vlan)#show conf
!
interface Vlan 4
  no ip address
  tagged Port-channel 1

Dell(conf-if-vlan)#end

Dell#show vlan

Codes: * - Default VLAN, G - GVRP VLANs

  NUM Status   Q  Ports
* 1   Inactive
  2   Active   T  Po1(So 0/0-1)
               T  Gi 3/0
  3   Active   T  Po1(So 0/0-1)
               T  Gi 3/1
  4   Active   T  Po1(So 0/0-1)
Dell#
```

When you remove a tagged interface from a VLAN (using the `no tagged interface` command), it remains tagged only if it is a tagged interface in another VLAN. If the tagged interface is removed from the only VLAN to which it belongs, the interface is placed in the Default VLAN as an untagged interface.

Moving Untagged Interfaces

To move untagged interfaces from the Default VLAN to another VLAN, use the following commands.

1. Access INTERFACE VLAN mode of the VLAN to which you want to assign the interface.
   CONFIGURATION mode

   `interface vlan *vlan-id*`
2. Configure an interface as untagged.

INTERFACE mode

`untagged interface`

This command is available only in VLAN interfaces.

Move an Untagged Interface to Another VLAN

The `no untagged` *`interface`* command removes the untagged interface from a port-based VLAN and places the interface in the Default VLAN. You cannot use the `no untagged` *`interface`* command in the Default VLAN. The following example shows the steps and commands to move an untagged interface from the Default VLAN to another VLAN.

To determine interface status, use the `show vlan` command. Interface (gi 3/2) is untagged and in the Default VLAN (vlan 1). In a port-based VLAN (vlan 4), use the `untagged` command to add the interface to that VLAN. The `show vlan` command output displays the interface's changed status (gi 3/2). Because the Default VLAN no longer contains any interfaces, it is listed as inactive.

```
Dell#show vlan

Codes: * - Default VLAN, G - GVRP VLANs

  NUM  Status  Q  Ports
* 1    Active  U  Gi 3/2
  2    Active  T  Po1(So 0/0-1)
               T  Gi 3/0
  3    Active  T  Po1(So 0/0-1)
               T  Gi 3/1
  4    Inactive
Dell#conf
Dell(conf)#int vlan 4
Dell(conf-if-vlan)#untagged gi 3/2
Dell(conf-if-vlan)#show config
!
interface Vlan 4
  no ip address
  untagged GigabitEthernet 3/2
Dell(conf-if-vlan)#end

Dell#show vlan

Codes: * - Default VLAN, G - GVRP VLANs

  NUM  Status   Q  Ports
* 1    Inactive
  2    Active   T  Po1(So 0/0-1)
                T  Gi 3/0
  3    Active   T  Po1(So 0/0-1)
                T  Gi 3/1
  4    Active   U  Gi 3/2
Dell#
```

The only way to remove an interface from the Default VLAN is to place the interface in Default mode by using the `no switchport` command in INTERFACE mode.

Assigning an IP Address to a VLAN

VLANs are a Layer 2 feature. For two physical interfaces on different VLANs to communicate, you must assign an IP address to the VLANs to route traffic between the two interfaces.
The `shutdown` command in INTERFACE mode does not affect Layer 2 traffic on the interface; the `shutdown` command only prevents Layer 3 traffic from traversing over the interface.

NOTE: You cannot assign an IP address to the Default VLAN (VLAN 1). To assign another VLAN ID to the Default VLAN, use the `default vlan-id` *`vlan-id`* command.

In Dell Networking OS, you can place VLANs and other logical interfaces in Layer 3 mode to receive and send routed traffic. For more information, refer to Bulk Configuration.

To assign an IP address, use the following command.

- Configure an IP address and mask on the interface.

  INTERFACE mode

  ```
  ip address ip-address mask [secondary]
  ```

  - *`ip-address mask`* — Enter an address in dotted-decimal format (A.B.C.D) and the mask must be in slash format (/24).
  - `secondary` — This is the interface's backup IP address. You can configure up to eight secondary IP addresses.

Configuring Native VLANs

Traditionally, ports can be either untagged for membership to one VLAN or tagged for membership to multiple VLANs.
You must connect an untagged port to a VLAN-unaware station (one that does not understand VLAN tags), and you must connect a tagged port to a VLAN-aware station (one that generates and understands VLAN tags).

Native VLAN support breaks this barrier so that you can connect a port to both VLAN-aware and VLAN-unaware stations. Such ports are referred to as hybrid ports. Physical and port-channel interfaces may be hybrid ports.

Native VLAN is useful in deployments where a Layer 2 port can receive both tagged and untagged traffic on the same physical port. The classic example is connecting a voice-over-IP (VOIP) phone and a PC to the same port of the switch. The VOIP phone is configured to generate tagged packets (with VLAN = VOICE VLAN) and the attached PC generates untagged packets.

NOTE: When a hybrid port is untagged in a VLAN but it receives tagged traffic, all traffic is accepted.

NOTE: You cannot configure an existing switchport or port channel interface for Native VLAN. Interfaces must have no other Layer 2 or Layer 3 configurations when using the `portmode hybrid` command or a message similar to this displays: `% Error: Port is in Layer-2 mode Gi 5/6.`

To configure a port so that it can be a member of an untagged and tagged VLANs, use the following commands.

1. Remove any Layer 2 or Layer 3 configurations from the interface.

INTERFACE mode

2. Configure the interface for Hybrid mode.

   INTERFACE mode

   ```
   portmode hybrid
   ```
3. Configure the interface for Switchport mode.

   INTERFACE mode

   ```
   switchport
   ```
4. Add the interface to a tagged or untagged VLAN.

   VLAN INTERFACE mode

   ```
   [tagged | untagged]
   ```

Enabling Null VLAN as the Default VLAN

In a Carrier Ethernet for Metro Service environment, service providers who perform frequent reconfigurations for customers with changing requirements occasionally enable multiple interfaces, each connected to a different customer, before the interfaces are fully configured.

This presents a vulnerability because both interfaces are initially placed in the native VLAN, VLAN 1, and for that period customers are able to access each other's networks. Dell Networking OS has a Null VLAN to eliminate this vulnerability. When you enable the Null VLAN, all ports are placed into it by default, so even if you activate the physical ports of multiple customers, no traffic is allowed to traverse the links until each port is place in another VLAN.

To enable Null VLAN, use the following command.

- Disable the default VLAN, so that all ports belong to the Null VLAN until configured as a member of another VLAN.

  CONFIGURATION mode

  ```
  default-vlan disable
  ```

  Default: the default VLAN is enabled (`no default-vlan disable`).

51

VLT Proxy Gateway

You can configure a proxy gateway in VLT domains. A proxy gateway enables you to locally route the packets that are destined to a L3 endpoint in another VLT domain.

Proxy Gateway in VLT Domains

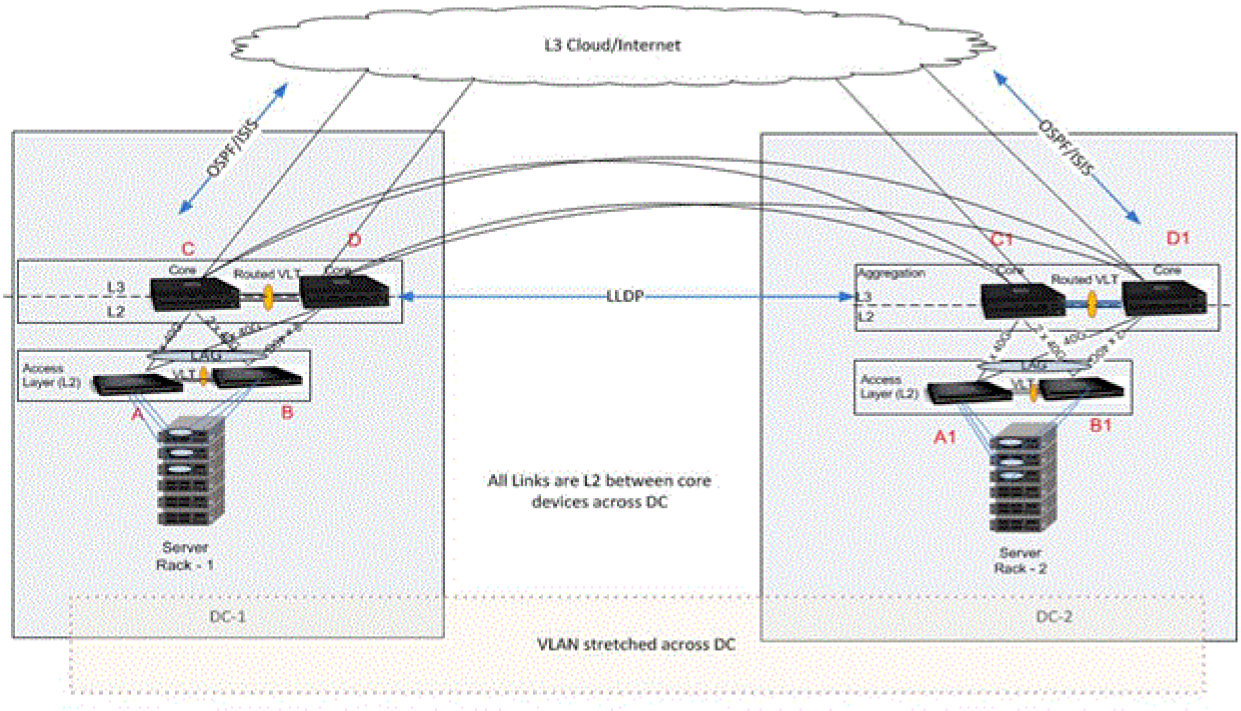
Using a proxy gateway, the VLT peers in a domain can route the L3 packets destined for VLT peers in another domain as long as they have L3 reachability of these IP destinations.

A proxy gateway in a VLT domain provides the following benefits:

- Avoid sub-optimal routing of packets by a VLT domain when packets are destined to the endpoint in another VLT domain.
- Provide resiliency if a VLT peer goes down by performing proxy routing for the peer's DA MAC in another VLT domain.

A typical scenario is virtual movement of servers across data centers. Virtual movement enables live migration of running virtual machines from one host to another without a downtime. Consider a square VLT connecting two data centers. If a VM, say VM1 on Server Rack 1 has C as its default gateway and VM1 performs a virtual movement to Server Rack 2 with no change in default gateway, then L3 packets destined for C can be routed either C1 or D1 locally. This behavior is achieved by installing the local_DA entries for C/D in C1/D1 so that the packets for C/D could have a hit at C1/D1 and routed locally.

In the following figure, server racks, named Rack 1 and Rack 2 are part of data centers, named DC1 and DC2 respectively. Rack 1 is connected to devices A and B in Layer 2. Similarly, Rack 2 is connected to devices A and B in Layer 2. A VLT LAG is present between A and B. A and B are connected to core routers, C and D. VLT routing is present between C and D. C1 and D1 are Layer 3 routers in DC2 and they are connected with core routers, C and D. The core or Layer 3 routers are then part of a Layer 3 cloud.


L3 Cloud/Internet
OSPF/ISIS
OSPF/ISIS
C
D
C1
D1
Core
Routed VLT
Core
L3
L2
LLDP
Aggregation
Core
Routed VLT
Core
L3
L2
LAG
Access Layer (L2)
VLT
LAG
Access Layer (L2)
VLT
A
B
A1
B1
All Links are L2 between core devices across DC
Server Rack - 1
Server Rack - 2
DC-1
DC-2
VLAN stretched across DC

Topology 1

When the routing table across DCs is not symmetrical, there is a possibility of a routing miss by a DC that do not have the route for the L3 traffic. Since routing protocols will enabled and both the DC's comes in same subnet there will not be route asymmetry dynamically. But if static route is configured on one DC and not on the other, it will result is asymmetry. Proxy routing can still be achieved locally by configuring a static route or default gateway.

Guidelines for Configuring a VLT Proxy Gateway

Keep the following points in mind when you configure this functionality:

1. Proxy gateway is supported only for VLT i.e. across VLT domain.

2. The current design will not handle the asymmetric VLAN configuration scenarios such as same VLAN configured with L2 mode on one VLT domain and L3 mode on the other VLT domain. It is always required to configure same mode for the VLAN's across the VLT domain.

3. VLAN symmetry within a VLT domain is also to be maintained.

4. The connection between DC's can only be a VLT.

5. Trace route across DC's may possibly show extra hops.

6. Route symmetry has to be maintained across the VLT domains to ensure no traffic drops.

7. If the port-channel specified in the proxy-gateway command is not a VLT LAG then the configuration is rejected by CLI. The VLT LAG cannot be changed to a legacy LAG when it is part of proxy-gateway.

8. LLDP port channel interface can’t be changed to legacy lag when proxy gateway is enabled.

9.“vlt-peer-mac transmit” is recommended only for square VLT without any diagonal links.

10. VRRP and IPv6 routing is not supported now.

11. With the existing hardware capabilities, only 512 my_station_tcam entries can be supported.

12. PVLAN not supported

13. After VM Motion, it’s expected that VM Host will send GARP in term, host previous VLT Domain will have mac movement points to newer VLT Domain

14. After station move, it is expected that if host send TTL1 packet destined to its gateway i.e previous Vlt Node, the packet may get dropped.

15. After station move, it’s expected that if host first PING its gateway (i.e previous VLT node) it would results in 40 to 60% success rate considering it take long path

Configuring a VLT Proxy Gateway

The VLT proxy gateway feature can be configured in a VLT domain context using the cli command `proxy-gateway LLDP`. You enter the proxy-gateway Configuration mode when you enter this command. The port-channel interface of the square VLT link on which LLDP packets are to be sent is specified by `peer-domain-link port-channel` command.

On a proxy gateway interface configuration corresponding to LLDP, LLDP sets TLV flags on the interfaces for receiving and transmitting private TLV packets. After defining these organizational TLV settings, LLDP encodes the proxy gateway TLVs based on the organizational TLVs for transmitting to the peer. If you specify the `no proxy gateway LLDP interface` command, LLDP stops transmitting and receiving proxy gateway TLV packets on the specified interfaces. However, other TLVs are not affected. Because of the timing defined in the LLDP configuration and the operational state, LLDP periodically sends or receives packets. However, the local DA updates may not be able to reach the destination on time. From the interfaces on which proxy gateway LLDP is enabled, LLDP decodes TLV packets from the remote LLDP by using the new organizational TLV.

The following requirements must be satisfied for LLDP proxy gateway to function correctly:

- As LLDP is direct link protocol, Data Centers must be directly connected.
- LLDP has a limited TLV size. As a result, information that is carried by this new TLV is limited to only one or two MACs.
- Proper configuration and physical setup must be ensured on all related systems.

LLDP organizational TLV for proxy gateway

A new organizational TLV is defined for this purpose:

- LLDP will define an organizationally specific TLV (type 127) with organizationally unique identifier (0x0001E8) and organizationally defined subtype (0x01) for sending or receiving this information.
- LLDP will use existing infrastructure but just adding this new TLV, and send and receive only on configured ports

- There are only a couple of MACs for each unit to be transmitted so that all current active MACs can definitely be carried on the newly defined TLV.
- This TLV is recognizable only by FTOS devices with this feature support. Other device will ignore this field and should still be able to process other standard TLVs.

The LLDP organizational TLV passes local DA information to peer VLT domain devices so they can act as proxy gateway. Two configurations are sent to LLDP to enable this feature:

- Global proxy gateway LLDP configuration to enable this feature
- Interface proxy gateway LLDP configuration to enable/disable proxy-gateway LLDP TLV on particular interfaces

- The interface is typically a port channel which connects to a remote VLT domain.
- The new proxy gateway TLV will be carried on the physical links under the port channel only
- There should be at least one link connects to each unit of the VLT domain

The configuration is complete if these two configurations are applied, with the following prerequisities:

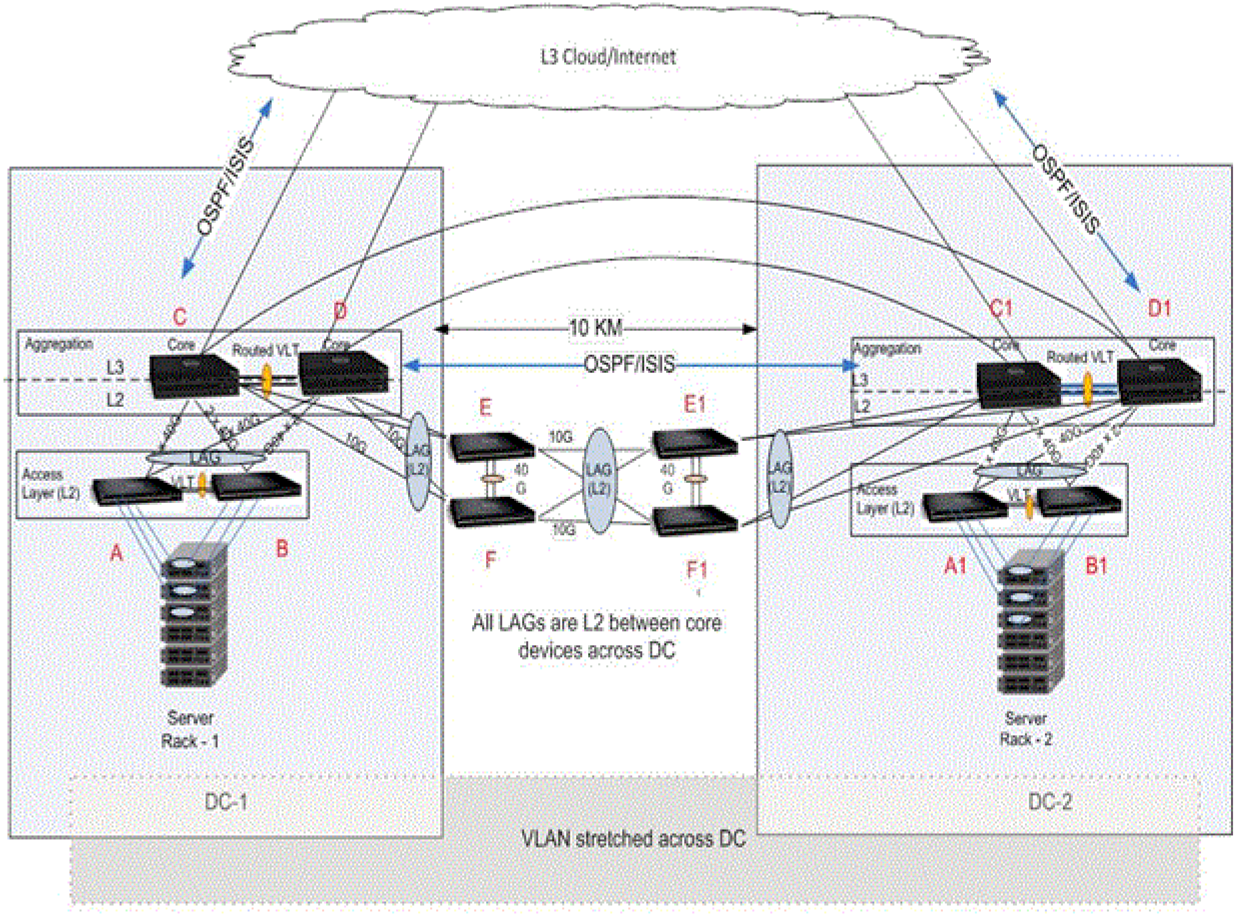
- LLDP must be globally enabled;
- There should not be any interface level LLDP disable CLI on the interfaces configured for proxy gateway, and both transmission and reception must be enabled;
- Both units of the remote VLT domain must be connected by the port channel member.
- If more than one port is connected to a unit of the remote VLT domain, it has to be completed by the time proxy gateway LLDP is enabled
- no other conflict configuration (i.e., no static proxy gateway configuration)

This feature might not operate properly if one of the following conditions is true:

- Any proxy gateway configuration or LLDP configuration is not working
- LLDP packets fail to reach to remote VLT domain devices (due to system down, rebooting, port down, physical link connection)

Sample Configuration Scenario for VLT Proxy Gateway

1. Assume the following topology with C1/D1 part of VLT domain 1 and C2/D2 part of VLT domain 2. This will undergo sub-optimal routing. The following figure illustrates a sample VLT Proxy gateway scenario.


L3 Cloud/Internet
OSPF/ISIS
OSPF/ISIS
C
D
C1
D1
Aggregation
Core
Routed VLT
Core
L3
L2
10 KM
OSPF/ISIS
E
E1
F
F1
10G
40 G
LAG (L2)
Access Layer (L2)
VLT
LAG
A
B
A1
B1
All LAGs are L2 between core devices across DC
Server Rack - 1
Server Rack - 2
DC-1
DC-2
VLAN stretched across DC

Topology 2

2. Trace route across VLT domains may show extra hops.
3. IP route symmetry must be maintained across the VLT domains. Assume if the route to a destination is not available at C2, though the packet hits the MY_STATION_TCAM and routing is enabled for that VLAN, if there is no entry for that prefix in the routing table it will dropped to CPU. By default, all route miss packets are given to CPU. To avoid this static entry must be configured.
4. There could be L3 frames received out-of-order at the L3 cloud, when a MAC is removed and added back. This could happen when proxy-routing and sub-optimal routing intersperse each other.
5. This feature is not supported for IPv6.
6. ICL shut – Assume ICL between C1 and D1 is shut and if D1 is secondary VLT then one half of the inter DC link goes down. After vm motion, If a packet reaches D2 with the destination MAC address of D1, it may be dropped. This behaviour is applicable only in the LLDP configuration; In static configuration, the packet will be forwarded.
7. Any L3 packet that was originally should have been switched across domains, when gets a hit at my_station_tcam (because of this feature) and routed, will have a TTL decrement as expected.

8. Packet duplication – Assume exclude-vlan (say VLAN 10) is configured on C2/D2 for C1's MAC. If packets for VLAN 10 with C1's MAC get a hit at C2, they will be switched to both D2 (via ICL) and C1 via inter DC link. This could lead to packet duplication. So, if C1's MAC is learnt at C2 then the packet would not have flooded (to D2) and only switched to C1 and thus avoided packet duplication.

Configuring a Static VLT Proxy Gateway

You can configure a proxy gateway in VLT domains. A proxy gateway enables you to locally route the packets that are destined to L3 endpoint of the other VLT domain. This functionality is supported on the S4810, S4820T, S6000, Z9000, and MXL platforms.

To configure the static proxy gateway, perform the following:

1. Enable VLT on a switch, then configure a VLT domain and enter VLT-domain configuration mode.

   CONFIGURATION mode

   ```
   Dell(conf)#vlt domain domain-id
   ```

2. Configure the static proxy gateway

   VLT DOMAIN mode

   ```
   Dell(conf-vlt-domain)#proxy-gateway static
   ```

3. You can configure the remote MAC address of a VLT peer for a static proxy gateway and exclude a VLAN or a range of VLANs from proxy routing. This parameter is for a static proxy gateway configuration.

   VLT DOMAIN PROXY GW STATIC mode

   ```
   Dell(conf-vlt-domain-proxy-gw-static)#remote-mac-address mac-address
   exclude-vlan vlan-range
   ```

4. Display the VLT proxy gateway configuration.

   EXEC mode

   ```
   Dell#show vlt-proxy-gateway
   ```

Configuring an LLDP VLT Proxy Gateway

You can configure a proxy gateway in a VLT domain to locally route packets destined to a L3 endpoint in another VLT domain.

To configure an LLDP proxy gateway:

1. Enable VLT on a switch, then configure a VLT domain and enter VLT-domain configuration mode.

   CONFIGURATION mode

   ```
   Dell(conf)#vlt domain domain-id
   ```

2. Configure the LLDP proxy gateway

   VLT DOMAIN mode

   ```
   Dell(conf-vlt-domain)#proxy-gateway lldp
   ```

3. You can configure the port channel interface for an LLDP proxy gateway and exclude a VLAN or a range of VLANs from proxy routing. This parameter is for an LLDP proxy gateway configuration.

VLT DOMAIN PROXY GW LLDP mode

```
Dell(conf-vlt-domain-proxy-gw-lldp)#peer-domain-link port-channel interface
exclude-vlan vlan-range
```

4. Display the VLT proxy gateway configuration.
   EXEC mode

   ```
   Dell#show vlt-proxy-gateway
   ```

52

Virtual Link Trunking (VLT)

Virtual link trunking (VLT) is supported on the S6000 platform.

Overview

VLT allows physical links between two chassis to appear as a single virtual link to the network core or other switches such as Edge, Access, or top-of-rack (ToR).

VLT reduces the role of spanning tree protocols (STPs) by allowing link aggregation group (LAG) terminations on two separate distribution or core switches, and by supporting a loop-free topology. (To prevent the initial loop that may occur prior to VLT being established, use a spanning tree protocol. After VLT is established, you may use rapid spanning tree protocol (RSTP) to prevent loops from forming with new links that are incorrectly connected and outside the VLT domain.)

VLT provides Layer 2 multipathing, creating redundancy through increased bandwidth, enabling multiple parallel paths between nodes and load-balancing traffic where alternative paths exist.

Virtual link trunking offers the following benefits:

- Allows a single device to use a LAG across two upstream devices.
- Eliminates STP-blocked ports.
- Provides a loop-free topology.
- Uses all available uplink bandwidth.
- Provides fast convergence if either the link or a device fails.
- Optimized forwarding with virtual router redundancy protocol (VRRP).
- Provides link-level resiliency.
- Assures high availability.

CAUTION: Dell Networking does not recommend enabling Stacking and VLT simultaneously. If you enable both features at the same time, unexpected behavior occurs.

As shown in the following example, VLT presents a single logical Layer 2 domain from the perspective of attached devices that have a virtual link trunk terminating on separate chassis in the VLT domain. However, the two VLT chassis are independent Layer2/Layer3 (L2/L3) switches for devices in the upstream network. L2/L3 control plane protocols and system management features function normally in VLT mode. Features such as VRRP and internet group management protocol (IGMP) snooping require state information coordinating between the two VLT chassis. IGMP and VLT configurations must be identical on both sides of the trunk to ensure the same behavior on both sides.

The following example shows VLT deployed on switches. The switches appear as a single virtual switch from the point of view of the switch or server supporting link aggregation control protocol (LACP).

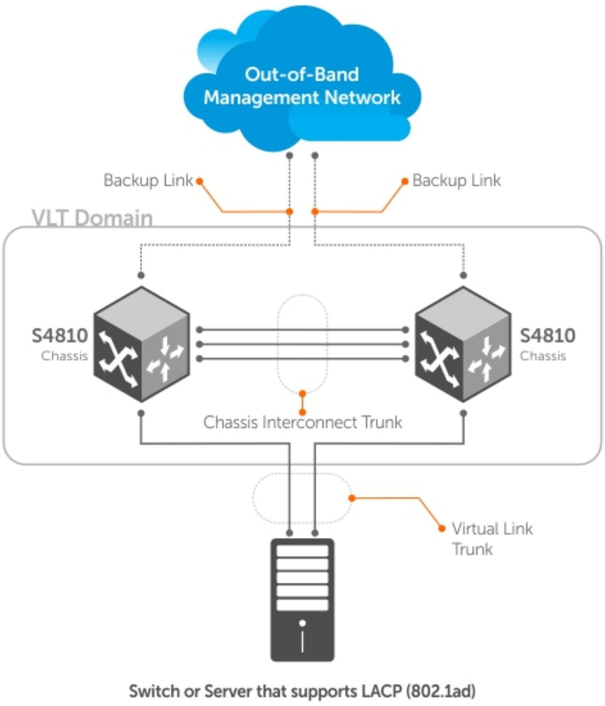

Out-of-Band Management Network
Backup Link
Backup Link
VLT Domain
S4810
Chassis
S4810
Chassis
Chassis Interconnect Trunk
Virtual Link Trunk
Switch or Server that supports LACP (802.1ad)

Figure 118. VLT on Switches

VLT on Core Switches

You can also deploy VLT on core switches.

Uplinks from servers to the access layer and from access layer to the aggregation layer are bundled in LAG groups with end-to-end Layer 2 multipathing. This set up requires "horizontal" stacking at the access layer and VLT at the aggregation layer such that all the uplinks from servers to access and access to aggregation are in Active-Active Load Sharing mode. This example provides the highest form of resiliency, scaling, and load balancing in data center switching networks.

The following example shows stacking at the access, VLT in aggregation, and Layer 3 at the core.

The aggregation layer is mostly in the L2/L3 switching/routing layer. For better resiliency in the aggregation, Dell Networking recommends running the internal gateway protocol (IGP) on the VLTi VLAN to synchronize the L3 routing table across the two nodes on a VLT system.

Enhanced VLT

An enhanced VLT (eVLT) configuration creates a port channel between two VLT domains by allowing two different VLT domains, using different VLT domain ID numbers, connected by a standard link aggregation control protocol (LACP) LAG to form a loop-free Layer 2 topology in the aggregation layer.

This configuration supports a maximum of four units, increasing the number of available ports and allowing for dual redundancy of the VLT. The following example shows how the core/aggregation port density in the Layer 2 topology is increased using eVLT. For inter-VLAN routing and other Layer 3 routing, you need a separate Layer 3 router.

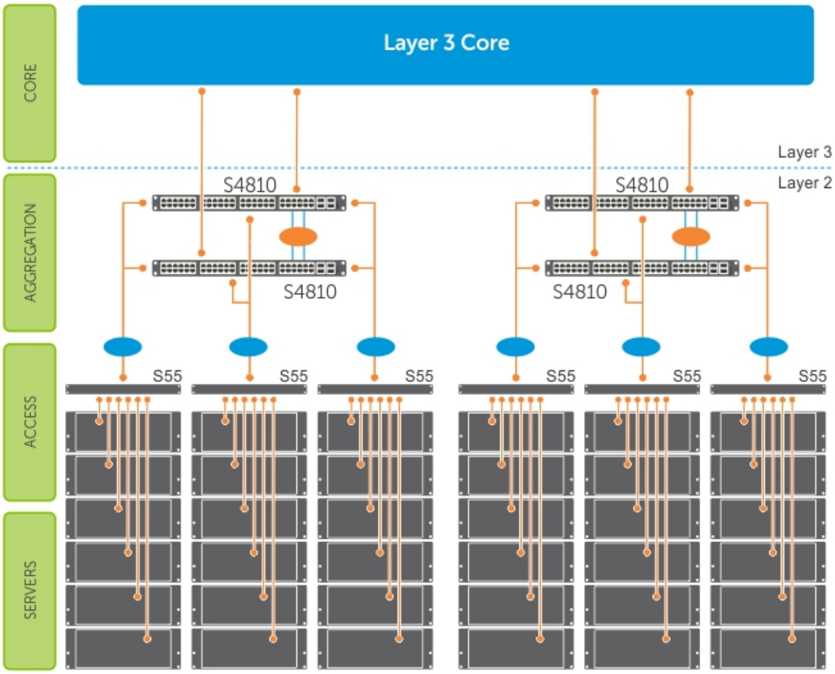


Figure 119. Enhanced VLT

VLT Terminology

The following are key VLT terms.

- **Virtual link trunk (VLT)** — The combined port channel between an attached device and the VLT peer switches.
- **VLT backup link** — The backup link monitors the vitality of VLT peer switches. The backup link sends configurable, periodic keep alive messages between the VLT peer switches.
- **VLT interconnect (VLTi)** — The link used to synchronize states between the VLT peer switches. Both ends must be on 10G or 40G interfaces.
- **VLT domain** — This domain includes both the VLT peer devices, VLT interconnect, and all of the port channels in the VLT connected to the attached devices. It is also associated to the configuration mode that you must use to assign VLT global parameters.
- **VLT peer device** — One of a pair of devices that are connected with the special port channel known as the VLT interconnect (VLTi).

VLT peer switches have independent management planes. A VLT interconnect between the VLT chassis maintains synchronization of L2/L3 control planes across the two VLT peer switches. The VLT interconnect uses either 10G or 40G user ports on the chassis.

A separate backup link maintains heartbeat messages across an out-of-band (OOB) management network. The backup link ensures that node failure conditions are correctly detected and are not confused with failures of the VLT interconnect. VLT ensures that local traffic on a chassis does not traverse the VLTi and takes the shortest path to the destination via directly attached links.

Configure Virtual Link Trunking

VLT requires that you enable the feature and then configure the same VLT domain, backup link, and VLT interconnect on both peer switches.

Important Points to Remember

- VLT port channel interfaces must be switch ports.
- If you include RSTP on the system, configure it before VLT. Refer to Configure Rapid Spanning Tree.
- Dell Networking strongly recommends that the VLTi (VLT interconnect) be a static LAG and that you disable LACP on the VLTi.
- Ensure that the spanning tree root bridge is at the Aggregation layer. If you enable RSTP on the VLT device, refer to RSTP and VLT for guidelines to avoid traffic loss.
- If you reboot both VLT peers in BMP mode and the VLT LAGs are static, the DHCP server reply to the DHCP discover offer may not be forwarded by the ToR to the correct node. To avoid this scenario, configure the VLT LAGs to the ToR and the ToR port channel to the VLT peers with LACP. If supported by the ToR, enable the `lacp-ungroup` feature on the ToR using the `lacp ungroup member-independent port-channel` command.
- If the `lacp-ungroup` feature is not supported on the ToR, reboot the VLT peers one at a time. After rebooting, verify that VLTi (ICL) is active before attempting DHCP connectivity.
- When you enable IGMP snooping on the VLT peers, ensure the value of the `delay-restore` command is not less than the query interval.
- When you enable Layer 3 routing protocols on VLT peers, make sure the delay-restore timer is set to a value that allows sufficient time for all routes to establish adjacency and exchange all the L3 routes between the VLT peers before you enable the VLT ports.
- Only use the `lacp ungroup member-independent` command if the system connects to nodes using bare metal provisioning (BMP) to upgrade or boot from the network.
- Ensure that you configure all port channels where LACP ungroup is applicable as hybrid ports and as untagged members of a VLAN. BMP uses untagged dynamic host configuration protocol (DHCP) packets to communicate with the DHCP server.
- If the DHCP server is located on the ToR and the VLTi (ICL) is down due to a failed link when a VLT node is rebooted in BMP mode, it is not able to reach the DHCP server, resulting in BMP failure.
- If the source is connected to an orphan (non-spanned, non-VLT) port in a VLT peer, the receiver is connected to a VLT (spanned) port-channel, and the VLT port-channel link between the VLT peer connected to the source and TOR is down, traffic is duplicated due to route inconsistency between peers. To avoid this scenario, Dell Networking recommends configuring both the source and the receiver on a spanned VLT VLAN.
- Bulk Sync happens only for Global IPv6 Neighbors; Link-local neighbor entries are not synced.
- If all of the following conditions are true, MAC addresses may not be synced correctly:
  - VLT peers use VLT interconnect (VLTi)
  - Sticky MAC is enabled on an orphan port in the primary or secondary peer
  - MACs are currently inactive

  If this scenario occurs, use the `clear mac-address-table sticky all` command on the primary or secondary peer to correctly sync the MAC addresses.
- If static ARP is enabled on only one VLT peer, entries may be overwritten during bulk sync.
- For multiple VLT LAGs configured on the same VLAN, if a host is learned on one VLT LAG and there is a station move between LAGs, the link local address redirects to the VLTi link on one of the peers. If this occurs, clear the link local address that is redirecting to the VLTi link.
- VLT Heartbeat is supported only on default VRFs.

- In a scenario where one hundred hosts are connected to a Peer1 on a non-VLT domain and traffic flows through Peer1 to Peer2; when you move these hosts from a non-VLT domain to a VLT domain and send ARP requests to Peer1, only half of these ARP requests reach Peer1, while the remaining half reach Peer2 (beacuse of LAG hashing). The reason for this behavior is that Peer1 ignores the ARP requests that it receives on VLTi (ICL) and updates only the ARP requests that it receives on the local VLT. As a result, the remaining ARP requests still points to the Non-VLT links and traffic does not reach half of the hosts. To mitigate this issue, ensure that you configure the following settings on both the Peers (Peer1 and Peer2): `arp learn-enable` and `mac-address-table station-move refresh-arp`.
- In a topology in which two VLT peer nodes that are connected by a VLTi link and are connected to a ToR switch using a VLT LAG interface, if you configure an egress IP ACL and apply it on the VLT LAG of both peers using the `deny ip any any` command, the traffic is permitted on the VLT LAG instead of being denied. The correct behavior of dropping the traffic on the VLT LAG occurs when VLT is up on both the peer nodes. However, if VLT goes down on one of the peers, traffic traverses through VLTi and the other peer switches it to the VLT LAG. Although egress ACL is applied on the VLT nodes to deny all traffic, this egress ACL does not deny the traffic (switching traffic is not denied owing to the egress IP ACL). You cannot use egress ACLs to deny traffic properly in such a VLT scenario.

Configuration Notes

When you configure VLT, the following conditions apply.

- VLT domain
  - A VLT domain supports two chassis members, which appear as a single logical device to network access devices connected to VLT ports through a port channel.
  - A VLT domain consists of the two core chassis, the interconnect trunk, backup link, and the LAG members connected to attached devices.
  - Each VLT domain has a unique MAC address that you create or VLT creates automatically.
  - ARP tables are synchronized between the VLT peer nodes.
  - VLT peer switches operate as separate chassis with independent control and data planes for devices attached on non-VLT ports.
  - One chassis in the VLT domain is assigned a primary role; the other chassis takes the secondary role. The primary and secondary roles are required for scenarios when connectivity between the chassis is lost. VLT assigns the primary chassis role according to the lowest MAC address. You can configure the primary role.
  - In a VLT domain, the peer switches must run the same Dell Networking OS software version.
  - Separately configure each VLT peer switch with the same VLT domain ID and the VLT version. If the system detects mismatches between VLT peer switches in the VLT domain ID or VLT version, the VLT Interconnect (VLTi) does not activate. To find the reason for the VLTi being down, use the `show vlt statistics` command to verify that there are mismatch errors, then use the `show vlt brief` command on each VLT peer to view the VLT version on the peer switch. If the VLT version is more than one release different from the current version in use, the VLTi does not activate.
  - The chassis members in a VLT domain support connection to orphan hosts and switches that are not connected to both switches in the VLT core.
- VLT interconnect (VLTi)
  - The VLT interconnect must consist of either 10G or 40G ports. A maximum of eight 10G or four 40G ports is supported. A combination of 10G and 40G ports is not supported.
  - A VLT interconnect over 1G ports is *not* supported.
  - The port channel must be in Default mode (not Switchport mode) to have VLTi recognize it.

  - The system automatically includes the required VLANs in VLTi. You do not need to manually select VLANs.
  - VLT peer switches operate as separate chassis with independent control and data planes for devices attached to non-VLT ports.
  - Port-channel link aggregation (LAG) across the ports in the VLT interconnect is required; individual ports are not supported. Dell Networking strongly recommends configuring a static LAG for VLTi.
  - The VLT interconnect synchronizes L2 and L3 control-plane information across the two chassis.
  - The VLT interconnect is used for data traffic only when there is a link failure that requires using VLTi in order for data packets to reach their final destination.
  - Unknown, multicast, and broadcast traffic can be flooded across the VLT interconnect.
  - MAC addresses for VLANs configured across VLT peer chassis are synchronized over the VLT interconnect on an egress port such as a VLT LAG. MAC addresses are the same on both VLT peer nodes.
  - ARP entries configured across the VLTi are the same on both VLT peer nodes.
  - If you shut down the port channel used in the VLT interconnect on a peer switch in a VLT domain in which you did not configure a backup link, the switch's role displays in the `show vlt brief` command output as Primary instead of Standalone.
  - When you change the default VLAN ID on a VLT peer switch, the VLT interconnect may flap.
  - In a VLT domain, the following software features are supported on VLTi: link layer discovery protocol (LLDP), flow control, port monitoring, jumbo frames, and data center bridging (DCB).
  - When you enable the VLTi link, the link between the VLT peer switches is established if the following configured information is true on both peer switches:
    - the VLT system MAC address matches.
    - the VLT unit-id is not identical.

    NOTE: If you configure the VLT system MAC address or VLT unit-id on only one of the VLT peer switches, the link between the VLT peer switches is not established. Each VLT peer switch must be correctly configured to establish the link between the peers.
  - If the link between the VLT peer switches is established, changing the VLT system MAC address or the VLT unit-id causes the link between the VLT peer switches to become disabled. However, removing the VLT system MAC address or the VLT unit-id may disable the VLT ports if you happen to configure the unit ID or system MAC address on only one VLT peer at any time.
  - If the link between VLT peer switches is established, any change to the VLT system MAC address or unit-id fails if the changes made create a mismatch by causing the VLT unit-ID to be the same on both peers and/or the VLT system MAC address does not match on both peers.
  - If you replace a VLT peer node, preconfigure the switch with the VLT system MAC address, unit-id, and other VLT parameters before connecting it to the existing VLT peer switch using the VLTi connection.
  - If the size of the MTU for VLTi members is less than 1496 bytes, MAC addresses may not be synced. Dell Networking recommends retaining the default MTU allocation (1554 bytes) for VLTi members.

- VLT backup link
  - In the backup link between peer switches, heartbeat messages are exchanged between the two chassis for health checks. The default time interval between heartbeat messages over the backup link is 1 second. You can configure this interval. The range is from 1 to 5 seconds. DSCP marking on heartbeat messages is CS6.
  - In order that the chassis backup link does not share the same physical path as the interconnect trunk, Dell Networking recommends using the management ports on the chassis and traverse an out-of-band management network. The backup link can use user ports, but not the same ports the interconnect trunk uses.

  - The chassis backup link does not carry control plane information or data traffic. Its use is restricted to health checks only.
- Virtual link trunks (VLTs) between access devices and VLT peer switches
  - To connect servers and access switches with VLT peer switches, you use a VLT port channel, as shown in Overview. Up to 48 port-channels are supported; up to eight member links are supported in each port channel between the VLT domain and an access device.
  - To connect servers and access switches with VLT peer switches, you use a VLT port channel, as shown in Overview. Up to 96 port-channels are supported; up to eight member links are supported in each port channel between the VLT domain and an access device.
  - The discovery protocol running between VLT peers automatically generates the ID number of the port channel that connects an access device and a VLT switch. The discovery protocol uses LACP properties to identify connectivity to a common client device and automatically generates a VLT number for port channels on VLT peers that connects to the device. The discovery protocol requires that an attached device always runs LACP over the port-channel interface.
  - VLT provides a loop-free topology for port channels with endpoints on different chassis in the VLT domain.
  - VLT uses shortest path routing so that traffic destined to hosts via directly attached links on a chassis does not traverse the chassis-interconnect link.
  - VLT allows multiple active parallel paths from access switches to VLT chassis.
  - VLT supports port-channel links with LACP between access switches and VLT peer switches. Dell Networking recommends using static port channels on VLTi.
  - If VLTi connectivity with a peer is lost but the VLT backup connectivity indicates that the peer is still alive, the VLT ports on the Secondary peer are orphaned and are shut down.
    * In one possible topology, a switch uses the BMP feature to receive its IP address, configuration files, and boot image from a DHCP server that connects to the switch through the VLT domain. In the port-channel used by the switch to connect to the VLT domain, configure the port interfaces on each VLT peer as hybrid ports before adding them to the port channel (refer to Connecting a VLT Domain to an Attached Access Device (Switch or Server)). To configure a port in Hybrid mode so that it can carry untagged, single-tagged, and double-tagged traffic, use the `portmode hybrid` command in Interface Configuration mode as described in Configuring Native VLANs.
    * For example, if the DHCP server is on the ToR and VLTi (ICL) is down (due to either an unavailable peer or a link failure), whether you configured the VLT LAG as static or LACP, when a single VLT peer is rebooted in BMP mode, it cannot reach the DHCP server, resulting in BMP failure.
- Software features supported on VLT port-channels
  - In a VLT domain, the following software features are supported on VLT port-channels: 802.1p, ingress and egress ACLs, BGP, DHCP relay, IS-IS, OSPF, active-active PIM-SM, PIM-SSM, VRRP, Layer 3 VLANs, LLDP, flow control, port monitoring, jumbo frames, IGMP snooping, sFlow, ingress and egress ACLs, and Layer 2 control protocols RSTP only).

    NOTE: PVST+ passthrough is supported in a VLT domain. PVST+ BPDUs does not result in an interface shutdown. PVST+ BPDUs for a nondefault VLAN is flooded out as any other L2 multicast packet. On a default VLAN, RTSP is part of the PVST+ topology in that specific VLAN (default VLAN).
  - For detailed information about how to use VRRP in a VLT domain, refer to the following *VLT and VRRP interoperability* section.
  - For information about configuring IGMP Snooping in a VLT domain, refer to VLT and IGMP Snooping.
  - All system management protocols are supported on VLT ports, including SNMP, RMON, AAA, ACL, DNS, FTP, SSH, Syslog, NTP, RADIUS, SCP, TACACS+, Telnet, and LLDP.
  - Enable Layer 3 VLAN connectivity VLT peers by configuring a VLAN network interface for the same VLAN on both switches.

  - Dell Networking does not recommend enabling peer-routing if the CAM is full. To enable peer-routing, a minimum of two local DA spaces for wild card functionality are required.

- Software features supported on VLT physical ports
  - In a VLT domain, the following software features are supported on VLT physical ports: 802.1p, LLDP, flow control, IPv6 dynamic routing, port monitoring, and jumbo frames.
- Software features not supported with VLT
  - In a VLT domain, the following software features are supported on non-VLT ports: 802.1x, DHCP snooping, FRRP, ingress and egress QOS.
- VLT and VRRP interoperability
  - In a VLT domain, VRRP interoperates with virtual link trunks that carry traffic to and from access devices (refer to Overview). The VLT peers belong to the same VRRP group and are assigned master and backup roles. Each peer actively forwards L3 traffic, reducing the traffic flow over the VLT interconnect.
  - VRRP elects the router with the highest priority as the master in the VRRP group. To ensure VRRP operation in a VLT domain, configure VRRP group priority on each VLT peer so that a peer is either the master or backup for all VRRP groups configured on its interfaces. For more information, refer to Setting VRRP Group (Virtual Router) Priority.
  - To verify that a VLT peer is consistently configured for either the master or backup role in all VRRP groups, use the `show vrrp` command on each peer.
  - Also configure the same L3 routing (static and dynamic) on each peer so that the L3 reachability and routing tables are identical on both VLT peers. Both the VRRP master and backup peers must be able to locally forward L3 traffic in the same way.
  - In a VLT domain, although both VLT peers actively participate in L3 forwarding as the VRRP master or backup router, the `show vrrp` command output displays one peer as master and the other peer as backup.
- Failure scenarios
  - On a link failover, when a VLT port channel fails, the traffic destined for that VLT port channel is redirected to the VLTi to avoid flooding.
  - When a VLT switch determines that a VLT port channel has failed (and that no other local port channels are available), the peer with the failed port channel notifies the remote peer that it no longer has an active port channel for a link. The remote peer then enables data forwarding across the interconnect trunk for packets that would otherwise have been forwarded over the failed port channel. This mechanism ensures reachability and provides loop management. If the VLT interconnect fails, the VLT software on the primary switch checks the status of the remote peer using the backup link. If the remote peer is up, the secondary switch disables all VLT ports on its device to prevent loops.
  - If all ports in the VLT interconnect fail, or if the messaging infrastructure fails to communicate across the interconnect trunk, the VLT management system uses the backup link interface to determine whether the failure is a link-level failure or whether the remote peer has failed entirely. If the remote peer is still alive (heartbeat messages are still being received), the VLT secondary switch disables its VLT port channels. If keepalive messages from the peer are not being received, the peer continues to forward traffic, assuming that it is the last device available in the network. In either case, after recovery of the peer link or reestablishment of message forwarding across the interconnect trunk, the two VLT peers resynchronize any MAC addresses learned while communication was interrupted and the VLT system continues normal data forwarding.
  - If the primary chassis fails, the secondary chassis takes on the operational role of the primary.
- The SNMP MIB reports VLT statistics.

Primary and Secondary VLT Peers

Primary and Secondary VLT Peers are supported on the platform.

To prevent issues when connectivity between peers is lost, you can designate Primary and Secondary roles for VLT peers . You can elect or configure the Primary Peer. By default, the peer with the lowest MAC address is selected as the Primary Peer. You can configure another peer as the Primary Peer using the VLT `domain` *`domain-id`* `role priority` *`priority-value`* command.

If the VLTi link fails, the status of the remote VLT Primary Peer is checked using the backup link. If the remote VLT Primary Peer is available, the Secondary Peer disables all VLT ports to prevent loops.

If all ports in the VLTi link fail or if the communication between VLTi links fails, VLT checks the backup link to determine the cause of the failure. If the failed peer can still transmit heartbeat messages, the Secondary Peer disables all VLT member ports and any Layer 3 interfaces attached to the VLAN associated with the VLT domain. If heartbeat messages are not received, the Secondary Peer forwards traffic assumes the role of the Primary Peer. If the original Primary Peer is restored, the VLT peer reassigned as the Primary Peer retains this role and the other peer must be reassigned as a Secondary Peer. Peer role changes are reported as SNMP traps.

RSTP and VLT

VLT provides loop-free redundant topologies and does not require RSTP.

RSTP can cause temporary port state blocking and may cause topology changes after link or node failures. Spanning tree topology changes are distributed to the entire layer 2 network, which can cause a network-wide flush of learned MAC and ARP addresses, requiring these addresses to be re-learned. However, enabling RSTP can detect potential loops caused by non-system issues such as cabling errors or incorrect configurations. To minimize possible topology changes after link or node failure, RSTP is useful for potential loop detection. Configure RSTP using the following specifications.

The following recommendations help you avoid these issues and the associated traffic loss caused by using RSTP when you enable VLT on both VLT peers:

- Configure any ports at the edge of the spanning tree's operating domain as edge ports, which are directly connected to end stations or server racks. Disable RSTP on ports connected directly to Layer 3-only routers not running STP or configure them as edge ports.
- Ensure that the primary VLT node is the root bridge and the secondary VLT peer node has the second-best bridge ID in the network. If the primary VLT peer node fails, the secondary VLT peer node becomes the root bridge, avoiding problems with spanning tree port state changes that occur when a VLT node fails or recovers.
- Even with this configuration, if the node has non-VLT ports using RSTP that you did not configure as edge ports and are connected to other Layer 2 switches, spanning tree topology changes are still detected after VLT node recovery. To avoid this scenario, ensure that you configure any non-VLT ports as edge ports or disable RSTP.

VLT Bandwidth Monitoring

When bandwidth usage of the VLTi (ICL) exceeds 80%, a syslog error message (shown in the following message) and an SNMP trap are generated.

```
%STKUNIT0-M:CP %VLTMGR-6-VLT-LAG-ICL: Overall Bandwidth utilization of VLT-ICL-
LAG (port-channel 25)
crosses threshold. Bandwidth usage (80 )
```

When the bandwidth usage drops below the 80% threshold, the system generates another syslog message (shown in the following message) and an SNMP trap.

```
%STKUNIT0-M:CP %VLTMGR-6-VLT-LAG-ICL: Overall Bandwidth utilization of VLT-ICL-
LAG (port-channel 25)
reaches below threshold. Bandwidth usage (74 )VLT show remote port channel
status
```

VLT and IGMP Snooping

When configuring IGMP Snooping with VLT, ensure the configurations on both sides of the VLT trunk are identical to get the same behavior on both sides of the trunk.

When you configure IGMP snooping on a VLT node, the dynamically learned groups and multicast router ports are automatically learned on the VLT peer node.

VLT IPv6

VLT IPv6 is supported on the platform.

The following features have been enhanced to support IPv6:

- **VLT Sync** — Entries learned on the VLT interface are synced on both VLT peers.
- **Non-VLT Sync** — Entries learned on non-VLT interfaces are synced on both VLT peers.
- **Tunneling** — Control information is associated with tunnel traffic so that the appropriate VLT peer can mirror the ingress port as the VLT interface rather than pointing to the VLT peer's VLTi link.
- **Statistics and Counters** — Statistical and counter information displays IPv6 information when applicable.
- **Heartbeat** — You can configure an IPv4 or IPv6 address as a backup link destination. You cannot use an IPv4 and an IPv6 address simultaneously.

VLT Port Delayed Restoration

When a VLT node boots up, if the VLT ports have been previously saved in the start-up configuration, they are not immediately enabled.

To ensure MAC and ARP entries from the VLT per node are downloaded to the newly enabled VLT node, the system allows time for the VLT ports on the new node to be enabled and begin receiving traffic.

The `delay-restore` feature waits for all saved configurations to be applied, then starts a configurable timer. After the timer expires, the VLT ports are enabled one-by-one in a controlled manner. The delay between bringing up each VLT port-channel is proportional to the number of physical members in the port-channel. The default is 90 seconds.

To change the duration of the configurable timer, use the `delay-restore` command.

If you enable IGMP snooping, IGMP queries are also sent out on the VLT ports at this time allowing any receivers to respond to the queries and update the multicast table on the new node.

This delay in bringing up the VLT ports also applies when the VLTi link recovers from a failure that caused the VLT ports on the secondary VLT peer node to be disabled.

PIM-Sparse Mode Support on VLT

The designated router functionality of the PIM Sparse-Mode multicast protocol is supported on VLT peer switches for multicast sources and receivers that are connected to VLT ports.

VLT peer switches can act as a last-hop router for IGMP receivers and as a first-hop router for multicast sources.

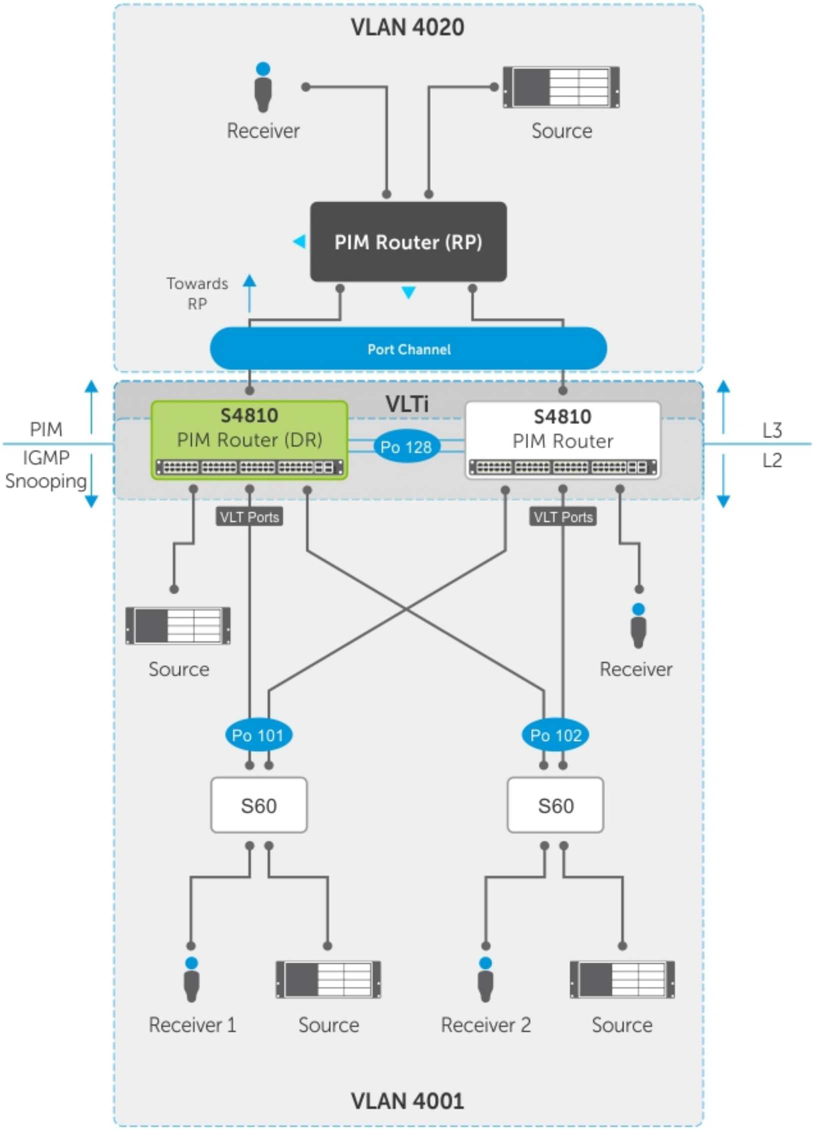

VLAN 4020
Receiver
Source
PIM Router (RP)
Towards RP
Port Channel
VLTi
S4810
PIM Router (DR)
Po 128
S4810
PIM Router
PIM
IGMP Snooping
L3
L2
VLT Ports
VLT Ports
Source
Receiver
Po 101
Po 102
S60
S60
Receiver 1
Source
Receiver 2
Source
VLAN 4001

Figure 120. PIM-Sparse Mode Support on VLT

On each VLAN where the VLT peer nodes act as the first hop or last hop routers, one of the VLT peer nodes is elected as the PIM designated router. If you configured IGMP snooping along with PIM on the VLT VLANs, you must configure VLTi as the static multicast router port on both VLT peer switches. This ensures that for first hop routers, the packets from the source are redirected to the designated router

(DR) if they are incorrectly hashed. In addition to being first-hop or last -hop routers, the peer node can also act as an intermediate router.

On a VLT-enabled PIM router, if any PIM neighbor is reachable through a Spanned Layer 3 (L3) VLAN interface, this must be the **only** PIM-enabled interface to reach that neighbor. A Spanned L3 VLAN is any L3 VLAN configured on both peers in a VLT domain. This does not apply to server-side L2 VLT ports because they do not connect to any PIM routers. These VLT ports can be members of multiple PIM-enabled L3 VLANs for compatibility with IGMP.

To route traffic to and from the multicast source and receiver, enable PIM on the L3 side connected to the PIM router using the `ip pim sparse-mode` command.

Each VLT peer runs its own PIM protocol independently of other VLT peers. To ensure the PIM protocol states or multicast routing information base (MRIB) on the VLT peers are synced, if the incoming interface (IIF) and outgoing interface (OIF) are Spanned, the multicast route table is synced between the VLT peers.

To verify the PIM neighbors on the VLT VLAN and on the multicast port, use the `show ip pim neighbor`, `show ip igmp snooping mrouter`, and `show running config` commands.

You cannot configure VLT peer nodes as rendezvous points, but you can connect PIM routers to VLT ports.

If the VLT node elected as the designated router fails and you enable VLT Multicast Routing, multicast routes are synced to the other peer for traffic forwarding to ensure minimal traffic loss. If you did not enable VLT Multicast Routing, traffic loss occurs until the other VLT peer is selected as the DR.

VLT Routing

VLT routing is supported on the platform.

Layer 2 protocols from the ToR to the server are intra-rack and inter-rack. No spanning tree is required, but interoperability with spanning trees at the aggregation layer is supported. Communication is active-active, with no blocked links. MAC tables are synchronized between VLT nodes for bridging and you can enable IGMP snooping.

Because VLT ports are Layer 2 ports and not IP interfaces, VLT Unicast and VLT Multicast routing protocols do not operate directly on VLT ports. You must add the VLT ports as a member of one or more VLANs and assign IP addresses to these VLANs. VLT Unicast and VLT Multicast routing protocols require VLAN IP interfaces for operation. Protocols such as BGP, ISIS, OSPF, and PIM are compatible with VLT Unicast Routing and VLT Multicast Routing.

Spanned VLANs

Any VLAN configured on both VLT peer nodes is referred to as a Spanned VLAN. The VLT Interconnect (VLTi) port is automatically added as a member of the Spanned VLAN. As a result, any adjacent router connected to at least one VLT node on a Spanned VLAN subnet is directly reachable from both VLT peer nodes at the routing level.

VLT Unicast Routing

VLT unicast routing is supported on the S6000 platform.

VLT unicast routing locally routes packets destined for the L3 endpoint of the VLT peer. This method avoids sub-optimal routing. Peer-routing syncs the MAC addresses of both VLT peers and requires two local DA entries in TCAM. In case a VLT node is down, a timer that allows you to configure the amount of time needed for peer recovery provides resiliency. You can enable VLT unicast across multiple configurations using VLT links. You can enable ECMP on VLT nodes using VLT unicast.

VLT unicast routing is supported on both IPv6/IPv4. To enable VLT unicast routing, both VLT peers must be in L3 mode. Static route and routing protocols such as RIP, OSPF, ISIS, and BGP are supported. However, point-to-point configuration is not supported. To enable VLT unicast, VLAN configuration must be symmetrical on both peers. You cannot configure the same VLAN as Layer 2 on one node and as Layer 3 on the other node. Configuration mismatches are logged in the syslog and display in the `show vlt mismatch` command output.

If you enable VLT unicast routing, the following actions occur:

- L3 routing is enabled on any new IP address / IPv6 address configured for a VLAN interface that is up.
- L3 routing is enabled on any VLAN with an admin state of up.

NOTE: If the CAM is full, do not enable peer-routing.

NOTE: The peer routing and peer-routing-timeout is applicable for both IPv6/ IPv4.

Configuring VLT Unicast

To enable and configure VLT unicast, follow these steps.

1. Enable VLT on a switch, then configure a VLT domain and enter VLT-domain configuration mode.
   CONFIGURATION mode

   ```
   vlt domain domain-id
   ```
2. Enable peer-routing.
   VLT DOMAIN mode

   ```
   peer-routing
   ```
3. Configure the peer-routing timeout.
   VLT DOMAIN mode

   ```
   peer-routing—timeout value
   ```

   `value`: Specify a value (in seconds) from 1 to 65535.

VLT Multicast Routing

VLT multicast routing is supported on the platform.

VLT Multicast Routing provides resiliency to multicast routed traffic during the multicast routing protocol convergence period after a VLT link or VLT peer fails using the least intrusive method (PIM) and does not alter current protocol behavior.

Unlike VLT Unicast Routing, a normal multicast routing protocol does not exchange multicast routes between VLT peers. When you enable VLT Multicast Routing, the multicast routing table is synced between the VLT peers. Only multicast routes configured with a Spanned VLAN IP as their IIF are synced between VLT peers. For multicast routes with a Spanned VLAN IIF, only OIFs configured with a Spanned VLAN IP interface are synced between VLT peers.

The advantages of syncing the multicast routes between VLT peers are:

- **VLT resiliency** — After a VLT link or peer failure, if the traffic hashes to the VLT peer, the traffic continues to be routed using multicast until the PIM protocol detects the failure and adjusts the multicast distribution tree.
- **Optimal routing** — The VLT peer that receives the incoming traffic can directly route traffic to all downstream routers connected on VLT ports.
- **Optimal VLTi forwarding** — Only one copy of the incoming multicast traffic is sent on the VLTi for routing or forwarding to any orphan ports, rather than forwarding all the routed copies.

Important Points to Remember

- You cannot configure a VLT node as a rendezvous point (RP), but any PIM-SM compatible VLT node can serve as a designated router (DR).
- You can only use one spanned VLAN from a PIM-enabled VLT node to an external neighboring PIM router.
- If you connect multiple spanned VLANs to a PIM neighbor, or if both spanned and non-spanned VLANs can access the PIM neighbor, ECMP can cause the PIM protocol running on each VLT peer node to choose a different VLAN or IP route to reach the PIM neighbor. This can result in issues with multicast route syncing between peers.
- Both VLT peers require symmetric Layer 2 and Layer 3 configurations on both VLT peers for any spanned VLAN.
- For optimal performance, configure the VLT VLAN routing metrics to prefer VLT VLAN interfaces over non-VLT VLAN interfaces.
- When using factory default settings on a new switch deployed as a VLT node, packet loss may occur due to the requirement that all ports must be open.
- ECMP is not compatible on VLT nodes using VLT multicast. You must use a single VLAN.

Configuring VLT Multicast

To enable and configure VLT multicast, follow these steps.

1. Enable VLT on a switch, then configure a VLT domain and enter VLT-domain configuration mode.

   CONFIGURATION mode

   ```
   vlt domain domain-id
   ```

2. Enable peer-routing.

   VLT DOMAIN mode

   ```
   peer-routing
   ```

3. Configure the multicast peer-routing timeout.

   VLT DOMAIN mode

   ```
   multicast peer-routing—timeout value
   ```

   `value`: Specify a value (in seconds) from 1 to 1200.

4. Configure a PIM-SM compatible VLT node as a designated router (DR). For more information, refer to Configuring a Designated Router.
5. Configure a PIM-enabled external neighboring router as a rendezvous point (RP). For more information, refer to Configuring a Static Rendezvous Point.
6. Configure the VLT VLAN routing metrics to prefer VLT VLAN interfaces over non-VLT VLAN interfaces. For more information, refer to Classify Traffic.
7. Configure symmetrical Layer 2 and Layer 3 configurations on both VLT peers for any spanned VLAN.

Non-VLT ARP Sync

Synchronization for non-ARP routing table entries is supported on the platform.

ARP entries (including ND entries) learned on other ports are synced with the VLT peer to support station move scenarios.

NOTE: ARP entries learned on non-VLT, non-spanned VLANs are not synced with VLT peers.

RSTP Configuration

RSTP is supported in a VLT domain.

Before you configure VLT on peer switches, configure RSTP in the network. RSTP is required for initial loop prevention during the VLT startup phase. You may also use RSTP for loop prevention in the network outside of the VLT port channel. For information about how to configure RSTP, Rapid Spanning Tree Protocol (RSTP).

Run RSTP on both VLT peer switches. The primary VLT peer controls the RSTP states, such as forwarding and blocking, on both the primary and secondary peers. Dell Networking recommends configuring the primary VLT peer as the RSTP primary root device and configuring the secondary VLT peer as the RSTP secondary root device.

BPDUs use the MAC address of the primary VLT peer as the RSTP bridge ID in the designated bridge ID field. The primary VLT peer sends these BPDUs on VLT interfaces connected to access devices. The MAC address for a VLT domain is automatically selected on the peer switches when you create the domain (refer to Enabling VLT and Creating a VLT Domain).

Configure both ends of the VLT interconnect trunk with identical RSTP configurations. When you enable VLT, the `show spanning-tree rstp brief` command output displays VLT information (refer to Verifying a VLT Configuration).

Preventing Forwarding Loops in a VLT Domain

During the bootup of VLT peer switches, a forwarding loop may occur until the VLT configurations are applied on each switch and the primary/secondary roles are determined.
To prevent the interfaces in the VLT interconnect trunk and RSTP-enabled VLT ports from entering a Forwarding state and creating a traffic loop in a VLT domain, take the following steps.

1. Configure RSTP in the core network and on each peer switch as described in Rapid Spanning Tree Protocol (RSTP).

   Disabling RSTP on one VLT peer may result in a VLT domain failure.
2. Enable RSTP on each peer switch.

   PROTOCOL SPANNING TREE RSTP mode

   ```
   no disable
   ```
3. Configure each peer switch with a unique bridge priority.

   PROTOCOL SPANNING TREE RSTP mode

   ```
   bridge-priority
   ```

Sample RSTP Configuration

The following is a sample of an RSTP configuration.

Using the example shown in the Overview section as a sample VLT topology, the primary VLT switch sends BPDUs to an access device (switch or server) with its own RSTP bridge ID. BPDUs generated by an RSTP-enabled access device are only processed by the primary VLT switch. The secondary VLT switch tunnels the BPDUs that it receives to the primary VLT switch over the VLT interconnect. Only the primary VLT switch determines the RSTP roles and states on VLT ports and ensures that the VLT interconnect link is never blocked.

In the case of a primary VLT switch failure, the secondary switch starts sending BPDUs with its own bridge ID and inherits all the port states from the last synchronization with the primary switch. An access device never detects the change in primary/secondary roles and does not see it as a topology change.

The following examples show the RSTP configuration that you must perform on each peer switch to prevent forwarding loops.

Configure RSTP on VLT Peers to Prevent Forwarding Loops (VLT Peer 1)

```
Dell_VLTpeer1(conf)#protocol spanning-tree rstp
Dell_VLTpeer1(conf-rstp)#no disable
Dell_VLTpeer1(conf-rstp)#bridge-priority 4096
```

Configure RSTP on VLT Peers to Prevent Forwarding Loops (VLT Peer 2)

```
Dell_VLTpeer2(conf)#protocol spanning-tree rstp
Dell_VLTpeer2(conf-rstp)#no disable
Dell_VLTpeer2(conf-rstp)#bridge-priority 0
```

Configuring VLT

To configure VLT, use the following procedure.
Prerequisites: Before you begin, make sure that both VLT peer switches are running the same Dell Networking OS version and are configured for RSTP as described in RSTP Configuration. For VRRP operation, ensure that you configure VRRP groups and L3 routing on each VLT peer as described in *VLT and VRRP interoperability* in the Configuration Notes section.

1. Configure the VLT interconnect for the VLT domain. The primary and secondary switch roles in the VLT domain are automatically assigned after you configure both sides of the VLTi.

   NOTE: If you use a third-party ToR unit, to avoid potential problems if you reboot the VLT peers, Dell recommends using static LAGs on the VLTi between VLT peers.
2. Enable VLT and create a VLT domain ID. VLT automatically selects a system MAC address.
3. Configure a backup link for the VLT domain.
4. (Optional) Manually reconfigure the default VLT settings, such as the MAC address and VLT primary/secondary roles.
5. Connect the peer switches in a VLT domain to an attached access device (switch or server).

Configuring a VLT Interconnect

To configure a VLT interconnect, follow these steps.

1. Configure the port channel for the VLT interconnect on a VLT switch and enter interface configuration mode.

   CONFIGURATION mode

```
interface port-channel id-number
```

Enter the same port-channel number configured with the `peer-link port-channel` command as described in Enabling VLT and Creating a VLT Domain.

NOTE: To be included in the VLTi, the port channel must be in Default mode (`no switchport` or VLAN assigned).

2. Remove an IP address from the interface.

   INTERFACE PORT-CHANNEL mode

   ```
   no ip address
   ```

3. Add one or more port interfaces to the port channel.

   INTERFACE PORT-CHANNEL mode

   ```
   channel-member interface
   ```

   `interface`: specify one of the following interface types:

   - 1-Gigabit Ethernet: Enter `gigabitethernet` *`slot/port`*.
   - 10-Gigabit Ethernet: Enter `tengigabitethernet` *`slot/port`*.

4. Ensure that the port channel is active.

   INTERFACE PORT-CHANNEL mode

   ```
   no shutdown
   ```

5. Repeat Steps 1 to 4 on the VLT peer switch to configure the VLT interconnect.

Enabling VLT and Creating a VLT Domain

To enable VLT and create a VLT domain, use the following steps.

1. Enable VLT on a switch, then configure a VLT domain and enter VLT-domain configuration mode.

   CONFIGURATION mode

   ```
   vlt domain domain-id
   ```

   The domain ID range is from 1 to 1000.

   Configure the same domain ID on the peer switch to allow for common peering. VLT uses the domain ID to automatically create a VLT MAC address for the domain. If you do not configure the system explicitly, the system mac-address of the primary will be the VLT MAC address for the domain.

   To disable VLT, use the `no vlt domain` command.

   NOTE: Do not use MAC addresses such as “reserved” or “multicast.”

2. Configure the IP address of the management interface on the remote VLT peer to be used as the endpoint of the VLT backup link for sending out-of-band hello messages.

   VLT DOMAIN CONFIGURATION mode

   ```
   back-up destination ip-address [interval seconds]
   ```

You can optionally specify the time interval used to send hello messages. The range is from 1 to 5 seconds.

3. Configure the port channel to be used as the VLT interconnect between VLT peers in the domain.

   VLT DOMAIN CONFIGURATION mode

   ```
   peer-link port-channel id-number
   ```

4. (Optional) Prevent a possible loop during the bootup of a VLT peer switch or a device that accesses the VLT domain.

   CONFIGURATION mode

   ```
   lacp ungroup member-independent {vlt | port-channel port-channel-id}
   ```

   LACP on VLT ports (on a VLT switch or access device), which are members of the virtual link trunk, is not brought up until the VLT domain is recognized on the access device.

5. Repeat Steps 1 to 4 on the VLT peer switch to configure the IP address of this switch as the endpoint of the VLT backup link and to configure the same port channel for the VLT interconnect.

Configuring a VLT Backup Link

To configure a VLT backup link, use the following command.

1. Specify the management interface to be used for the backup link through an out-of-band management network.

   CONFIGURATION mode

   ```
   interface managementethernet slot/ port
   ```

   Enter the slot (0-1) and the port (0).

2. Configure an IPv4 address (A.B.C.D) or IPv6 address (X:X:X:X::X) and mask (/x) on the interface.

   MANAGEMENT INTERFACE mode

   ```
   {ip address ipv4-address/ mask | ipv6 address ipv6-address/ mask}
   ```

   This is the IP address to be configured on the VLT peer with the back-up destination command.

3. Ensure that the interface is active.

   MANAGEMENT INTERFACE mode

   ```
   no shutdown
   ```

4. Repeat Steps 1 to 3 on the VLT peer switch.

To set an amount of time, in seconds, to delay the system from restoring the VLT port, use the `delay-restore` command at any time. For more information, refer to VLT Port Delayed Restoration.

Configuring a VLT Port Delay Period

To configure a VLT port delay period, use the following commands.

1. Enter VLT-domain configuration mode for a specified VLT domain.

   CONFIGURATION mode

   ```
   vlt domain domain-id
   ```

The range of domain IDs from 1 to 1000.

2. Enter an amount of time, in seconds, to delay the restoration of the VLT ports after the system is rebooted.

   CONFIGURATION mode

   ```
   delay-restore delay-restore-time
   ```

   The range is from 1 to 1200.

   The default is **90 seconds**.

Reconfiguring the Default VLT Settings (Optional)

To reconfigure the default VLT settings, use the following commands.

1. Enter VLT-domain configuration mode for a specified VLT domain.

   CONFIGURATION mode

   ```
   vlt domain domain-id
   ```

   The range of domain IDs is from 1 to 1000.
2. (Optional) After you configure the VLT domain on each peer switch on both sides of the interconnect trunk, by default, Dell Networking OS elects a primary and secondary VLT peer device.

   VLT DOMAIN CONFIGURATION mode

   ```
   primary-priority value
   ```

   To reconfigure the primary role of VLT peer switches, use the `primary-priority` command. To configure the primary role on a VLT peer, enter a lower value than the priority value of the remote peer.

   The priority values are from 1 to 65535. The default is **32768**.
3. (Optional) When you create a VLT domain on a switch, Dell Networking OS automatically creates a VLT-system MAC address used for internal system operations.

   VLT DOMAIN CONFIGURATION mode

   ```
   system-mac mac-address mac-address
   ```

   To explicitly configure the default MAC address for the domain by entering a new MAC address, use the `system-mac` command. The format is aaaa.bbbb.cccc.

   Also, reconfigure the same MAC address on the VLT peer switch.

   Use this command to minimize the time required for the VLT system to synchronize the default MAC address of the VLT domain on both peer switches when one peer switch reboots.
4. (Optional) When you create a VLT domain on a switch, Dell Networking OS automatically assigns a unique unit ID (0 or 1) to each peer switch.

   VLT DOMAIN CONFIGURATION mode

   ```
   unit-id {0 | 1}
   ```

   To explicitly configure the default values on each peer switch, use the `unit-id` command.

Configure a different unit ID (0 or 1) on each peer switch.

Unit IDs are used for internal system operations.

Use this command to minimize the time required for the VLT system to determine the unit ID assigned to each peer switch when one peer switch reboots.

Connecting a VLT Domain to an Attached Access Device (Switch or Server)

To connect a VLT domain to an attached access device, use the following commands.
On a VLT peer switch: To connect to an attached device, configure the same port channel ID number on each peer switch in the VLT domain.

1. Configure the same port channel to be used to connect to an attached device and enter interface configuration mode.

   CONFIGURATION mode

   ```
   interface port-channel id-number
   ```
2. Remove an IP address from the interface.

   INTERFACE PORT-CHANNEL mode

   ```
   no ip address
   ```
3. Place the interface in Layer 2 mode.

   INTERFACE PORT-CHANNEL mode

   ```
   switchport
   ```
4. Add one or more port interfaces to the port channel.

   INTERFACE PORT-CHANNEL mode

   ```
   channel-member interface
   ```

   `interface`: specify one of the following interface types:

   - 1-Gigabit Ethernet: enter `gigabitethernet` *`slot/port`*.
   - 10-Gigabit Ethernet: enter `tengigabitethernet` *`slot/port`*.
5. Ensure that the port channel is active.

   INTERFACE PORT-CHANNEL mode

   ```
   no shutdown
   ```
6. Associate the port channel to the corresponding port channel in the VLT peer for the VLT connection to an attached device.

   INTERFACE PORT-CHANNEL mode

   ```
   vlt-peer-lag port-channel id-number
   ```

   The valid port-channel ID numbers are from 1 to 128.
7. Repeat Steps 1 to 6 on the VLT peer switch to configure the same port channel as part of the VLT domain.
8. **On an attached switch or server**: To connect to the VLT domain and add port channels to it, configure a port channel. For an example of how to verify the port-channel configuration, refer to VLT Sample Configuration.

To configure the VLAN where a VLT peer forwards received packets over the VLTi from an adjacent VLT peer that is down, use the `peer-down-vlan` parameter. When a VLT peer with BMP reboots, untagged DHCP discover packets are sent to the peer over the VLTi. Using this configuration ensures the DHCP discover packets are forwarded to the VLAN that has the DHCP server.

Configuring a VLT VLAN Peer-Down (Optional)

To configure a VLT VLAN peer-down, use the following commands.

1. Enter VLT-domain configuration mode for a specified VLT domain.

   CONFIGURATION mode

   ```
   vlt domain domain-id
   ```

   The range of domain IDs is from 1 to 1000.
2. Enter the port-channel number that acts as the interconnect trunk.

   VLT DOMAIN CONFIGURATION mode

   ```
   peer-link port-channel id-number
   ```

   The range is from 1 to 128.
3. Enter the VLAN ID number of the VLAN where the VLT forwards packets received on the VLTi from an adjacent peer that is down.

   VLT DOMAIN CONFIGURATION mode

   ```
   peer-down-vlan vlan interface number
   ```

   The range is from 1 to 4094.

Configuring Enhanced VLT (eVLT) (Optional)

To configure enhanced VLT (eVLT) between two VLT domains on your network, use the following procedure.
For a sample configuration, refer to eVLT Configuration Example. To set up the VLT domain, use the following commands.

1. Configure the port channel to be used for the VLT interconnect on a VLT switch and enter interface configuration mode.

   CONFIGURATION mode

   ```
   interface port-channel id-number
   ```

   Enter the same port-channel number configured with the `peer-link port-channel` command in the Enabling VLT and Creating a VLT Domain.
2. Add one or more port interfaces to the port channel.

   INTERFACE PORT-CHANNEL mode

   ```
   channel-member interface
   ```

   `interface`: specify one of the following interface types:

   - 1 Gigabit Ethernet: enter `gigabitethernet` *`slot/port`*.
   - 10 Gigabit Ethernet: enter `tengigabitethernet` *`slot/port`*.

3. Enter VLT-domain configuration mode for a specified VLT domain.

   CONFIGURATION mode

   ```
   vlt domain domain-id
   ```

   The range of domain IDs is from 1 to 1000.
4. Enter the port-channel number that acts as the interconnect trunk.

   VLT DOMAIN CONFIGURATION mode

   ```
   peer-link port-channel id-number
   ```

   The range is from 1 to 128.
5. Configure the IP address of the management interface on the remote VLT peer to be used as the endpoint of the VLT backup link for sending out-of-band hello messages.

   VLT DOMAIN CONFIGURATION mode

   ```
   back-up destination ip-address [interval seconds]
   ```

   You can optionally specify the time interval used to send hello messages. The range is from 1 to 5 seconds.
6. When you create a VLT domain on a switch, Dell Networking OS automatically creates a VLT-system MAC address used for internal system operations.

   VLT DOMAIN CONFIGURATION mode

   ```
   system-mac mac-address mac-address
   ```

   To explicitly configure the default MAC address for the domain by entering a new MAC address, use the `system-mac` command. The format is aaaa.bbbb.cccc.

   Also reconfigure the same MAC address on the VLT peer switch.

   Use this command to minimize the time required for the VLT system to synchronize the default MAC address of the VLT domain on both peer switches when one peer switch reboots.
7. When you create a VLT domain on a switch, Dell Networking OS automatically assigns a unique unit ID (0 or 1) to each peer switch.

   VLT DOMAIN CONFIGURATION mode

   ```
   unit-id {0 | 1}
   ```

   The unit IDs are used for internal system operations.

   To explicitly configure the default values on each peer switch, use the `unit-id` command.

   Configure a different unit ID (0 or 1) on each peer switch.

   Use this command to minimize the time required for the VLT system to determine the unit ID assigned to each peer switch when one peer switch reboots.
8. **Configure enhanced VLT.** Configure the port channel to be used for the VLT interconnect on a VLT switch and enter interface configuration mode.

   CONFIGURATION mode

```
interface port-channel id-number
```

Enter the same port-channel number configured with the `peer-link port-channel` command in the Enabling VLT and Creating a VLT Domain.

9. Place the interface in Layer 2 mode.

   INTERFACE PORT-CHANNEL mode

   ```
   switchport
   ```
10. Associate the port channel to the corresponding port channel in the VLT peer for the VLT connection to an attached device.

    INTERFACE PORT-CHANNEL mode

    ```
    vlt-peer-lag port-channel id-number
    ```

    Valid port-channel ID numbers are from 1 to 128.
11. Ensure that the port channel is active.

    INTERFACE PORT-CHANNEL mode

    ```
    no shutdown
    ```
12. **Add links to the eVLT port.** Configure a range of interfaces to bulk configure.

    CONFIGURATION mode

    ```
    interface range {port-channel id}
    ```
13. Enable LACP on the LAN port.

    INTERFACE mode

    ```
    port-channel-protocol lacp
    ```
14. Configure the LACP port channel mode.

    INTERFACE mode

    ```
    port-channel number mode [active]
    ```
15. Ensure that the interface is active.

    MANAGEMENT INTERFACE mode

    ```
    no shutdown
    ```
16. Repeat steps 1 through 15 for the VLT peer node in Domain 1.
17. Repeat steps 1 through 15 for the first VLT node in Domain 2.
18. Repeat steps 1 through 15 for the VLT peer node in Domain 2.

To verify the configuration of a VLT domain, use any of the `show` commands described in Verifying a VLT Configuration.

VLT Sample Configuration

To review a sample VLT configuration setup, study these steps.

1. Configure the VLT domain with the same ID in VLT peer 1 and VLT peer 2.

   VLT DOMAIN mode

```
vlt domain domain id
```

2. Configure the VLTi between VLT peer 1 and VLT peer 2.
3. You can configure LACP/static LAG between the peer units (not shown).

   CONFIGURATION mode

   ```
   interface port-channel port-channel id
   ```

   NOTE: To benefit from the protocol negotiations, Dell Networking recommends configuring VLTs used as facing hosts/switches with LACP. Ensure both peers use the same port channel ID.
4. Configure the peer-link port-channel in the VLT domains of each peer unit.

   INTERFACE PORTCHANNEL mode

   ```
   channel-member
   ```
5. Configure the backup link between the VLT peer units (shown in the following example).
6. Configure the peer 2 management ip/ interface ip for which connectivity is present in VLT peer 1.

   EXEC Privilege mode

   ```
   show running-config vlt
   ```
7. Configure the peer 1 management ip/ interface ip for which connectivity is present in VLT peer 1.

   EXEC mode or EXEC Privilege mode

   ```
   show interfaces interface
   ```
8. Configure the VLT links between VLT peer 1 and VLT peer 2 to the top of rack unit (shown in the following example).
9. Configure the static LAG/LACP between ports connected from VLT peer 1 and VLT peer 2 to the top of rack unit.

   EXEC Privilege mode

   ```
   show running-config entity
   ```
10. Configure the VLT peer link port channel id in VLT peer 1 and VLT peer 2.

    EXEC mode or EXEC Privilege mode

    ```
    show interfaces interface
    ```
11. In the top of rack unit, configure LACP in the physical ports.

    EXEC Privilege mode

    ```
    show running-config entity
    ```
12. Verify that VLT is running.

    EXEC mode

    `show vlt brief` or `show vlt detail`
13. Verify that the VLT LAG is running in both VLT peer units.

    EXEC mode or EXEC Privilege mode

    ```
    show interfaces interface
    ```

Example of Configuring VLT

In the following sample VLT configuration steps, VLT peer 1 is , VLT peer 2 is , and the ToR is S60-1.

NOTE: If you use a third-party ToR unit, Dell Networking recommends using static LAGs with VLT peers to avoid potential problems if you reboot the VLT peers.

Configure the VLT domain with the same ID in VLT peer 1 and VLT peer 2.

```
s4810-2(conf)#vlt domain 5
s4810-2(conf-vlt-domain)#

s4810-4(conf)#vlt domain 5
s4810-4(conf-vlt-domain)#
```

Configure the VLTi between VLT peer 1 and VLT peer 2.

1. You can configure the LACP/static LAG between the peer units (not shown).
2. Configure the peer-link port-channel in the VLT domains of each peer unit.

```
s4810-2(conf)#interface port-channel 1
s4810-2(conf-if-po-1)#channel-member TenGigabitEthernet 0/4-7
s4810-4(conf)#interface port-channel 1
s4810-4(conf-if-po-1)#channel-member TenGigabitEthernet 0/4-7
```

Configure the backup link between the VLT peer units.

1. Configure the peer 2 management ip/ interface ip for which connectivity is present in VLT peer 1.
2. Configure the peer 1 management ip/ interface ip for which connectivity is present in VLT peer 2.

```
s4810-2#show running-config vlt
!
vlt domain 5
  peer-link port-channel 1
  back-up destination 10.11.206.58
s4810-2#

s4810-2# show interfaces managementethernet 0/0
Internet address is 10.11.206.43/16

s4810-4#show running-config vlt
!
vlt domain 5
  peer-link port-channel 1
  back-up destination 10.11.206.43
s4810-4#
s4810-4#show running-config interface managementethernet 0/0
ip address 10.11.206.58/16
  no shutdown
```

Configure the VLT links between VLT peer 1 and VLT peer 2 to the Top of Rack unit. In the following example, port Te 0/40 in VLT peer 1 is connected to Te 0/48 of TOR and port Te 0/18 in VLT peer 2 is connected to Te 0/50 of TOR.

1. Configure the static LAG/LACP between the ports connected from VLT peer 1 and VLT peer 2 to the Top of Rack unit.
2. Configure the VLT peer link port channel id in VLT peer 1 and VLT peer 2.

3. In the Top of Rack unit, configure LACP in the physical ports (shown for VLT peer 1 only. Repeat steps for VLT peer 2. The bold `vlt-peer-lag port-channel 2` indicates that port-channel 2 is the port-channel id configured in VLT peer 2).

```
s4810-2#show running-config interface tengigabitethernet 0/40
!
interface TenGigabitEthernet 0/40
  no ip address
!
  port-channel-protocol LACP
    port-channel 2 mode active
  no shutdown
s4810-2#
configuring VLT peer lag in VLT
s4810-2#show running-config interface port-channel 2
!
interface Port-channel 2
  no ip address
  switchport
  vlt-peer-lag port-channel 2
  no shutdown
s4810-2#
s4810-2#show interfaces port-channel 2 brief
Codes: L - LACP Port-channel

   LAG  Mode  Status  Uptime    Ports
L  2    L2L3  up      03:33:14  Te 0/40 (Up)
s4810-2#
```

In the ToR unit, configure LACP on the physical ports.

```
s60-1#show running-config interface tengigabitethernet 0/48
!
interface TenGigabitEthernet 0/48
  no ip address
!
  port-channel-protocol LACP
    port-channel 100 mode active
  no shutdown

s60-1#show running-config interface tengigabitethernet 0/50
!
interface TenGigabitEthernet 0/50
no ip address
!
  port-channel-protocol LACP
    port-channel 100 mode active
  no shutdown
s60-1#

s60-1#show running-config interface port-channel 100
!
interface Port-channel 100
  no ip address
  switchport
  no shutdown
s60-1#
s60-1#show interfaces port-channel 100 brief
Codes: L - LACP Port-channel

   LAG  Mode  Status  Uptime    Ports
L  100  L2    up      03:33:48  Te 0/48 (Up)
```

```
                                Te 0/50 (Up)
s60-1#
```

Verify VLT is up. Verify that the VLTi (ICL) link, backup link connectivity (heartbeat status), and VLT peer link (peer chassis) are all up.

```
s4810-2#show vlt brief
  VLT Domain Brief
------------------
  Domain ID:                 5
  Role:                      Primary
  Role Priority:             32768
  ICL Link Status:           Up
  HeartBeat Status:          Up
  VLT Peer Status:           Up
  Local System MAC address:  00:01:e8:8c:4d:08
  Remote System MAC address: 00:01:e8:8c:4d:1c
s4810-2#

s4810-2#show vlt detail
Local LAG Id Peer LAG Id Local Status Active VLANs
------------ ----------- ------------ ------------
2            2           Up           1000-1199
```

Verify that the VLT LAG is up in both VLT peer units.

```
s4810-2#show interfaces port-channel 2 brief
Codes: L - LACP Port-channel

  LAG  Mode  Status  Uptime    Ports
L 2    L2L3  up      03:43:24  Te 0/40 (Up)
s4810-2#
s4810-4#show interfaces port-channel 2 brief
Codes: L - LACP Port-channel

  LAG  Mode  Status  Uptime    Ports
L 2    L2L3  up      03:33:31  Te 0/18 (Up)
s4810-4#
```

eVLT Configuration Example

The following example demonstrates the steps to configure enhanced VLT (eVLT) in a network.

In this example, you are configuring two domains. Domain 1 consists of Peer 1 and Peer 2; Domain 2 consists of Peer 3 and Peer 4, as shown in the following example.

In Domain 1, configure Peer 1 fist, then configure Peer 2. When that is complete, perform the same steps for the peer nodes in Domain 2. The interface used in this example is TenGigabitEthernet.

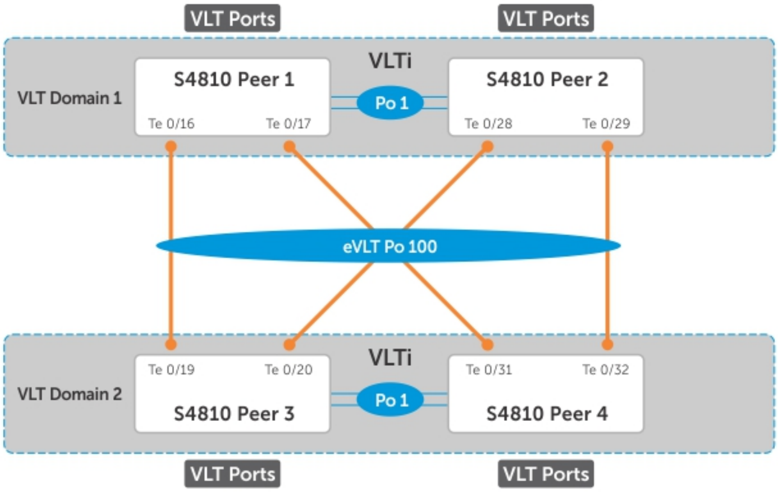

VLT Ports
VLT Ports
VLT Domain 1
S4810 Peer 1
VLTi
Po 1
S4810 Peer 2
Te 0/16
Te 0/17
Te 0/28
Te 0/29
eVLT Po 100
VLT Domain 2
Te 0/19
Te 0/20
VLTi
Te 0/31
Te 0/32
S4810 Peer 3
Po 1
S4810 Peer 4
VLT Ports
VLT Ports

Figure 121. eVLT Configuration Example

eVLT Configuration Step Examples

In Domain 1, configure the VLT domain and VLTi on Peer 1.

```
Domain_1_Peer1#configure
Domain_1_Peer1(conf)#interface port-channel 1
Domain_1_Peer1(conf-if-po-1)# channel-member TenGigabitEthernet 0/8-9

Domain_1_Peer1(conf)#vlt domain 1000
Domain_1_Peer1(conf-vlt-domain)# peer-link port-channel 1
Domain_1_Peer1(conf-vlt-domain)# back-up destination 10.16.130.11
Domain_1_Peer1(conf-vlt-domain)# system-mac mac-address 00:0a:00:0a:00:0a
Domain_1_Peer1(conf-vlt-domain)# unit-id 0
```

Configure eVLT on Peer 1.

```
Domain_1_Peer1(conf)#interface port-channel 100
Domain_1_Peer1(conf-if-po-100)# switchport
Domain_1_Peer1(conf-if-po-100)# vlt-peer-lag port-channel 100
Domain_1_Peer1(conf-if-po-100)# no shutdown
```

Add links to the eVLT port-channel on Peer 1.

```
Domain_1_Peer1(conf)#interface range tengigabitethernet 0/16 - 17
Domain_1_Peer1(conf-if-range-te-0/16-17)# port-channel-protocol LACP
Domain_1_Peer1(conf-if-range-te-0/16-17)# port-channel 100 mode active
Domain_1_Peer1(conf-if-range-te-0/16-17)# no shutdown
```

Next, configure the VLT domain and VLTi on Peer 2.

```
Domain_1_Peer2#configure
Domain_1_Peer2(conf)#interface port-channel 1
Domain_1_Peer2(conf-if-po-1)# channel-member TenGigabitEthernet 0/8-9

Domain_1_Peer2(conf) #vlt domain 1000
Domain_1_Peer2(conf-vlt-domain)# peer-link port-channel 1
```

```
Domain_1_Peer2(conf-vlt-domain)# back-up destination 10.16.130.12
Domain_1_Peer2(conf-vlt-domain)# system-mac mac-address 00:0a:00:0a:00:0a
Domain_1_Peer2(conf-vlt-domain)# unit-id 1
```

Configure eVLT on Peer 2.

```
Domain_1_Peer2(conf)#interface port-channel 100
Domain_1_Peer2(conf-if-po-100)# switchport
Domain_1_Peer2(conf-if-po-100)# vlt-peer-lag port-channel 100
Domain_1_Peer2(conf-if-po-100)# no shutdown
```

Add links to the eVLT port-channel on Peer 2.

```
Domain_1_Peer2(conf)#interface range tengigabitethernet 0/28 - 29
Domain_1_Peer2(conf-if-range-te-0/16-17)# port-channel-protocol LACP
Domain_1_Peer2(conf-if-range-te-0/16-17)# port-channel 100 mode active
Domain_1_Peer2(conf-if-range-te-0/16-17)# no shutdown
```

In Domain 2, configure the VLT domain and VLTi on Peer 3.

```
Domain_2_Peer3#configure
Domain_2_Peer3(conf)#interface port-channel 1
Domain_2_Peer3(conf-if-po-1)# channel-member TenGigabitEthernet 0/8-9
Domain_1_Peer3#no shutdown
Domain_2_Peer3(conf)#vlt domain 200
Domain_2_Peer3(conf-vlt-domain)# peer-link port-channel 1
Domain_2_Peer3(conf-vlt-domain)# back-up destination 10.18.130.11
Domain_2_Peer3(conf-vlt-domain)# system-mac mac-address 00:0b:00:0b:00:0b
Domain_2_Peer3(conf-vlt-domain)# unit-id 0
```

Configure eVLT on Peer 3.

```
Domain_2_Peer3(conf)#interface port-channel 100
Domain_2_Peer3(conf-if-po-100)# switchport
Domain_2_Peer3(conf-if-po-100)# vlt-peer-lag port-channel 100
Domain_2_Peer3(conf-if-po-100)# no shutdown
```

Add links to the eVLT port-channel on Peer 3.

```
Domain_2_Peer3(conf)#interface range tengigabitethernet 0/19 - 20
Domain_2_Peer3(conf-if-range-te-0/16-17)# port-channel-protocol LACP
Domain_2_Peer3(conf-if-range-te-0/16-17)# port-channel 100 mode active
Domain_2_Peer3(conf-if-range-te-0/16-17)# no shutdown
```

Next, configure the VLT domain and VLTi on Peer 4.

```
Domain_2_Peer4#configure
Domain_2_Peer4(conf)#interface port-channel 1
Domain_2_Peer4(conf-if-po-1)# channel-member TenGigabitEthernet 0/8-9
Domain_1_Peer4#no shutdown

Domain_2_Peer4(conf)#vlt domain 200
Domain_2_Peer4(conf-vlt-domain)# peer-link port-channel 1
Domain_2_Peer4(conf-vlt-domain)# back-up destination 10.18.130.12
Domain_2_Peer4(conf-vlt-domain)# system-mac mac-address 00:0b:00:0b:00:0b
Domain_2_Peer4(conf-vlt-domain)# unit-id 1
```

Configure eVLT on Peer 4.

```
Domain_2_Peer4(conf)#interface port-channel 100
Domain_2_Peer4(conf-if-po-100)# switchport
Domain_2_Peer4(conf-if-po-100)# vlt-peer-lag port-channel 100
Domain_2_Peer4(conf-if-po-100)# no shutdown
```

Add links to the eVLT port-channel on Peer 4.

```
Domain_2_Peer4(conf)#interface range tengigabitethernet 0/31 - 32
Domain_2_Peer4(conf-if-range-te-0/16-17)# port-channel-protocol LACP
Domain_2_Peer4(conf-if-range-te-0/16-17)# port-channel 100 mode active
Domain_2_Peer4(conf-if-range-te-0/16-17)# no shutdown
```

PIM-Sparse Mode Configuration Example

The following sample configuration shows how to configure the PIM Sparse mode designated router functionality on the VLT domain with two VLT port-channels that are members of VLAN 4001.

For more information, refer to PIM-Sparse Mode Support on VLT.

Examples of Configuring PIM-Sparse Mode

The following example shows how to enable PIM multicast routing on the VLT node globally.

```
VLT_Peer1(conf)#ip multicast-routing
```

The following example shows how to enable PIM on the VLT port VLANs.

```
VLT_Peer1(conf)#interface vlan 4001
VLT_Peer1(conf-if-vl-4001)#ip address 140.0.0.1/24
VLT_Peer1(conf-if-vl-4001)#ip pim sparse-mode
VLT_Peer1(conf-if-vl-4001)#tagged port-channel 101
VLT_Peer1(conf-if-vl-4001)#tagged port-channel 102
VLT_Peer1(conf-if-vl-4001)#no shutdown
VLT_Peer1(conf-if-vl-4001)#exit
```

The following example shows how to configure the VLTi port as a static multicast router port for the VLAN.

```
VLT_Peer1(conf)#interface vlan 4001
VLT_Peer1(conf-if-vl-4001)#ip igmp snooping mrouter interface port-channel 128
VLT_Peer1(conf-if-vl-4001)#exit
VLT_Peer1(conf)#end
```

The following example shows how to repeat these steps on VLT Peer Node 2.

```
VLT_Peer2(conf)#ip multicast-routing

VLT_Peer2(conf)#interface vlan 4001
VLT_Peer2(conf-if-vl-4001)#ip address 140.0.0.2/24
VLT_Peer2(conf-if-vl-4001)#ip pim sparse-mode
VLT_Peer2(conf-if-vl-4001)#tagged port-channel 101
VLT_Peer2(conf-if-vl-4001)#tagged port-channel 102
VLT_Peer2(conf-if-vl-4001)#no shutdown

VLT_Peer2(conf-if-vl-4001)#ip igmp snooping mrouter interface port-channel 128
VLT_Peer2(conf-if-vl-4001)#exit
VLT_Peer2(conf)#end
```

Verifying a VLT Configuration

To monitor the operation or verify the configuration of a VLT domain, use any of the following `show` commands on the primary and secondary VLT switches.

- Display information on backup link operation.

  EXEC mode

  ```
  show vlt backup-link
  ```
- Display general status information about VLT domains currently configured on the switch.

  EXEC mode

  ```
  show vlt brief
  ```
- Display detailed information about the VLT-domain configuration, including local and peer port-channel IDs, local VLT switch status, and number of active VLANs on each port channel.

  EXEC mode

  ```
  show vlt detail
  ```
- Display the VLT peer status, role of the local VLT switch, VLT system MAC address and system priority, and the MAC address and priority of the locally-attached VLT device.

  EXEC mode

  ```
  show vlt role
  ```
- Display the current configuration of all VLT domains or a specified group on the switch.

  EXEC mode

  ```
  show running-config vlt
  ```
- Display statistics on VLT operation.

  EXEC mode

  ```
  show vlt statistics
  ```
- Display the RSTP configuration on a VLT peer switch, including the status of port channels used in the VLT interconnect trunk and to connect to access devices.

  EXEC mode

  ```
  show spanning-tree rstp
  ```
- Display the current status of a port or port-channel interface used in the VLT domain.

  EXEC mode

  `show interfaces` *`interface`*

  – *`interface`*: specify one of the following interface types:

    * Fast Ethernet: enter `fastethernet` *`slot/port`*.
    * 1-Gigabit Ethernet: enter `gigabitethernet` *`slot/port`*.
    * 10-Gigabit Ethernet: enter `tengigabitethernet` *`slot/port`*.
    * Port channel: enter `port-channel {1-128}`.

Examples of the `show vlt` and `show spanning-tree rstp` Commands

The following example shows the `show vlt backup-link` command.

```
Dell_VLTpeer1# show vlt backup-link

VLT Backup Link
-----------------
Destination:                 10.11.200.18
Peer HeartBeat status:       Up
HeartBeat Timer Interval:    1
HeartBeat Timeout:           3
UDP Port:                    34998
HeartBeat Messages Sent:     1026
HeartBeat Messages Received: 1025

Dell_VLTpeer2# show vlt backup-link

VLT Backup Link
-----------------
Destination:                 10.11.200.20
Peer HeartBeat status:       Up
HeartBeat Timer Interval:    1
HeartBeat Timeout:           3
UDP Port:                    34998
HeartBeat Messages Sent:     1030
HeartBeat Messages Received: 1014
```

The following example shows the `show vlt brief` command.

```
Dell_VLTpeer1# show vlt brief
VLT Domain Brief
------------------
  Domain ID:                       1000
  Role:                            Secondary
  Role Priority:                   32768
  ICL Link Status:                 Up
  HeartBeat Status:                Up
  VLT Peer Status:                 Up
  Local Unit Id:                   0
  Version:                         5(1)
  Local System MAC address:        00:01:e8:8a:e9:70
  Remote System MAC address:       00:01:e8:8a:e7:e7
  Configured System MAC address: 00:0a:0a:01:01:0a
  Remote system version:           5(1)
  Delay-Restore timer:             90 seconds

Dell_VLTpeer2# show vlt brief
VLT Domain Brief
------------------
Domain ID:                         1000
  Role:                            Primary
  Role Priority:                   32768
  ICL Link Status:                 Up
  HeartBeat Status:                Up
  VLT Peer Status:                 Up
  Local Unit Id:                   1
  Version:                         5(1)
  Local System MAC address:        00:01:e8:8a:e7:e7
  Remote System MAC address:       00:01:e8:8a:e9:70
  Configured System MAC address: 00:0a:0a:01:01:0a
  Remote system version:           5(1)
  Delay-Restore timer:             90 seconds
```

The following example shows the `show vlt detail` command.

```
Dell_VLTpeer1# show vlt detail

Local LAG Id Peer LAG Id Local Status Peer Status Active VLANs
------------ ----------- ------------ ----------- -------------
100          100         UP            UP          10, 20, 30
127          2           UP            UP        20, 30

Dell_VLTpeer2# show vlt detail

Local LAG Id Peer LAG Id Local Status Peer Status Active VLANs
------------ ----------- ------------ ----------- -------------
2            127         UP           UP          20, 30
100          100         UP           UP          10, 20, 30
```

The following example shows the `show vlt role` command.

```
Dell_VLTpeer1# show vlt role

VLT Role
----------
VLT Role:                   Primary
System MAC address:         00:01:e8:8a:df:bc
System Role Priority:       32768
Local System MAC address:   00:01:e8:8a:df:bc
Local System Role Priority: 32768

Dell_VLTpeer2# show vlt role

VLT Role
----------
VLT Role:                   Secondary
System MAC address:         00:01:e8:8a:df:bc
System Role Priority:       32768
Local System MAC address:   00:01:e8:8a:df:e6
Local System Role Priority: 32768
```

The following example shows the `show running-config vlt` command.

```
Dell_VLTpeer1# show running-config vlt
!
vlt domain 30
  peer-link port-channel 60
  back-up destination 10.11.200.18

Dell_VLTpeer2# show running-config vlt
!
vlt domain 30
  peer-link port-channel 60
  back-up destination 10.11.200.20
```

The following example shows the `show vlt statistics` command.

```
Dell_VLTpeer1# show vlt statistics

VLT Statistics
----------------
HeartBeat Messages Sent:     987
HeartBeat Messages Received: 986
ICL Hello's Sent:            148
ICL Hello's Received:        98
```

```
Dell_VLTpeer2# show vlt statistics

VLT Statistics
----------------
HeartBeat Messages Sent:     994
HeartBeat Messages Received: 978
ICL Hello's Sent:            89
ICL Hello's Received:        89
```

The following example shows the `show spanning-tree rstp` command.

The bold section displays the RSTP state of port channels in the VLT domain. Port channel 100 is used in the VLT interconnect trunk (VLTi) to connect to VLT peer2. Port channels 110, 111, and 120 are used to connect to access switches or servers (vlt).

```
Dell_VLTpeer1# show spanning-tree rstp brief

Executing IEEE compatible Spanning Tree Protocol
Root ID Priority 0, Address 0001.e88a.dff8
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID Priority 4096, Address 0001.e88a.d656
Configured hello time 2, max age 20, forward delay 15

Interface                                  Designated
Name   PortID  Prio Cost  Sts Cost   Bridge ID PortID
---------- -------- ---- ------- --------- ------- ------------------
Po 1   128.2   128 200000 DIS       800   4096   0001.e88a.d656 128.2
Po 3   128.4   128 200000 DIS       800   4096   0001.e88a.d656 128.4
Po 4   128.5   128 200000 DIS       800   4096   0001.e88a.d656 128.5
Po 100 128.101 128 800    FWD(VLTi) 800   0      0001.e88a.dff8 128.101
Po 110 128.111 128 00     FWD(vlt)  800   4096   0001.e88a.d656 128.111
Po 111 128.112 128 200000 DIS(vlt)  800   4096   0001.e88a.d656 128.112
Po 120 128.121 128 2000   FWD(vlt)  800   4096   0001.e88a.d656 128.121

Dell_VLTpeer2# show spanning-tree rstp brief

Executing IEEE compatible Spanning Tree Protocol
Root ID Priority 0, Address 0001.e88a.dff8
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID Priority 0, Address 0001.e88a.dff8
We are the root
Configured hello time 2, max age 20, forward delay 15

Interface                                  Designated
Name   PortID  Prio Cost  Sts      Cost    Bridge ID PortID
---------- -------- ---- ------- -------- - ------- -------------
Po 1   128.2   128 200000 DIS      0   0   0001.e88a.dff8 128.2
Po 3   128.4   128 200000 DIS      0   0   0001.e88a.dff8 128.4
Po 4   128.5   128 200000 DIS      0   0   0001.e88a.dff8 128.5
Po 100 128.101 128 800    FWD(VLTi)0   0   0001.e88a.dff8 128.101
Po 110 128.111 128 00     FWD(vlt) 0   0   0001.e88a.dff8 128.111
Po 111 128.112 128 200000 DIS(vlt) 0   0   0001.e88a.dff8 128.112
Po 120 128.121 128 2000   FWD(vlt) 0   0   0001.e88a.dff8 128.121
```

Additional VLT Sample Configurations

To configure VLT, configure a backup link and interconnect trunk, create a VLT domain, configure a backup link and interconnect trunk, and connect the peer switches in a VLT domain to an attached access device (switch or server).

Review the following examples of VLT configurations.

Configuring Virtual Link Trunking (VLT Peer 1)

Enable VLT and create a VLT domain with a backup-link and interconnect trunk (VLTi).

```
Dell_VLTpeer1(conf)#vlt domain 999
Dell_VLTpeer1(conf-vlt-domain)#peer-link port-channel 100
Dell_VLTpeer1(conf-vlt-domain)#back-up destination 10.11.206.35
Dell_VLTpeer1(conf-vlt-domain)#exit
```

Configure the backup link.

```
Dell_VLTpeer1(conf)#interface ManagementEthernet 0/0
Dell_VLTpeer1(conf-if-ma-0/0)#ip address 10.11.206.23/
Dell_VLTpeer1(conf-if-ma-0/0)#no shutdown
Dell_VLTpeer1(conf-if-ma-0/0)#exit
```

Configure the VLT interconnect (VLTi).

```
Dell_VLTpeer1(conf)#interface port-channel 100
Dell_VLTpeer1(conf-if-po-100)#no ip address
Dell_VLTpeer1(conf-if-po-100)#channel-member fortyGigE 0/56,60
Dell_VLTpeer1(conf-if-po-100)#no shutdown
Dell_VLTpeer1(conf-if-po-100)#exit
```

Configure the port channel to an attached device.

```
Dell_VLTpeer1(conf)#interface port-channel 110
Dell_VLTpeer1(conf-if-po-110)#no ip address
Dell_VLTpeer1(conf-if-po-110)#switchport
Dell_VLTpeer1(conf-if-po-110)#channel-member fortyGigE 0/52
Dell_VLTpeer1(conf-if-po-110)#no shutdown
Dell_VLTpeer1(conf-if-po-110)#vlt-peer-lag port-channel 110
Dell_VLTpeer1(conf-if-po-110)#end
```

Verify that the port channels used in the VLT domain are assigned to the same VLAN.

```
Dell_VLTpeer1# show vlan id 10
Codes: * - Default VLAN, G - GVRP VLANs, P - Primary, C - Community, I -
Isolated
Q: U - Untagged, T - Tagged
   x - Dot1x untagged, X - Dot1x tagged
   G - GVRP tagged, M - Vlan-stack, H - Hyperpull tagged

   NUM Status Description Q Ports
   10  Active             U Po110(Fo 0/52)
                          T Po100(Fo 0/56,60)
```

Configuring Virtual Link Trunking (VLT Peer 2)

Enable VLT and create a VLT domain with a backup-link VLT interconnect (VLTi).

```
Dell_VLTpeer2(conf)#vlt domain 999
Dell_VLTpeer2(conf-vlt-domain)#peer-link port-channel 100
Dell_VLTpeer2(conf-vlt-domain)#back-up destination 10.11.206.23
Dell_VLTpeer2(conf-vlt-domain)#exit
```

Configure the backup link.

```
Dell_VLTpeer2(conf)#interface ManagementEthernet 0/0
Dell_VLTpeer2(conf-if-ma-0/0)#ip address 10.11.206.35/
Dell_VLTpeer2(conf-if-ma-0/0)#no shutdown
Dell_VLTpeer2(conf-if-ma-0/0)#exit
```

Configure the VLT interconnect (VLTi).

```
Dell_VLTpeer2(conf)#interface port-channel 100
Dell_VLTpeer2(conf-if-po-100)#no ip address
Dell_VLTpeer2(conf-if-po-100)#channel-member fortyGigE 0/46,50
Dell_VLTpeer2(conf-if-po-100)#no shutdown
Dell_VLTpeer2(conf-if-po-100)#exit
```

Configure the port channel to an attached device.

```
Dell_VLTpeer2(conf)#interface port-channel 110
Dell_VLTpeer2(conf-if-po-110)#no ip address
Dell_VLTpeer2(conf-if-po-110)#switchport
Dell_VLTpeer2(conf-if-po-110)#channel-member fortyGigE 0/48
Dell_VLTpeer2(conf-if-po-110)#no shutdown
Dell_VLTpeer2(conf-if-po-110)#vlt-peer-lag port-channel 110
Dell_VLTpeer2(conf-if-po-110)#end
```

Verify that the port channels used in the VLT domain are assigned to the same VLAN.

```
Dell_VLTpeer2# show vlan id 10
Codes: * - Default VLAN, G - GVRP VLANs, P - Primary, C - Community, I -
Isolated
Q: U - Untagged, T - Tagged
   x - Dot1x untagged, X - Dot1x tagged
   G - GVRP tagged, M - Vlan-stack, H - Hyperpull tagged

   NUM Status Description Q Ports
   10  Active             U Po110(Fo 0/48)
                          T Po100(Fo 0/46,50)
```

Verifying a Port-Channel Connection to a VLT Domain (From an Attached Access Switch)

On an access device, verify the port-channel connection to a VLT domain.

```
Dell_TORswitch(conf)# show running-config interface port-channel 11
!
interface Port-channel 11
no ip address
switchport
channel-member fortyGigE 1/18,22
no shutdown
```

Troubleshooting VLT

To help troubleshoot different VLT issues that may occur, use the following information.

NOTE: For information on VLT Failure mode timing and its impact, contact your Dell Networking representative.

Table 72. Troubleshooting VLT

Description	Behavior at Peer Up	Behavior During Run Time	Action to Take
Bandwidth monitoring	A syslog error message and an SNMP trap is generated when the VLTi bandwidth usage goes above the 80% threshold and when it drops below 80%.	A syslog error message and an SNMP trap is generated when the VLTi bandwidth usage goes above its threshold.	Depending on the traffic that is received, the traffic can be offloaded inVLTi.
Domain ID mismatch	The VLT peer does not boot up. The VLTi is forced to a down state. A syslog error message and an SNMP trap are generated.	The VLT peer does not boot up. The VLTi is forced to a down state. A syslog error message and an SNMP trap are generated.	Verify the domain ID matches on both VLT peers.
Dell Networking OS Version mismatch	A syslog error message is generated.	A syslog error message is generated.	Follow the correct upgrade procedure for the unit with the mismatched Dell Networking OS version.
Remote VLT port channel status	N/A	N/A	Use the `show vlt detail` and `show vlt brief` commands to view the VLT port channel status information.
Spanning tree mismatch at global level	All VLT port channels go down on both VLT peers. A syslog error message is generated.	No traffic is passed on the port channels. A one-time informational syslog message is generated.	During run time, a loop may occur as long as the mismatch lasts. To resolve, enable RSTP on both VLT peers.
Spanning tree mismatch at port level	A syslog error message is generated.	A one-time informational syslog message is generated.	Correct the spanning tree configuration on the ports.
System MAC mismatch	A syslog error message and an SNMP trap are generated.	A syslog error message and an SNMP trap are generated.	Verify that the unit ID of VLT peers is not the same on both units and

Description	Behavior at Peer Up	Behavior During Run Time	Action to Take
			that the MAC address is the same on both units.
Unit ID mismatch	The VLT peer does not boot up. The VLTi is forced to a down state. A syslog error message is generated.	The VLT peer does not boot up. The VLTi is forced to a down state. A syslog error message is generated.	Verify the unit ID is correct on both VLT peers. Unit ID numbers must be sequential on peer units; for example, if Peer 1 is unit ID “0”, Peer 2 unit ID must be “1’.
Version ID mismatch	A syslog error message and an SNMP trap are generated.	A syslog error message and an SNMP trap are generated.	Verify the Dell Networking OS software versions on the VLT peers is compatible. For more information, refer to the *Release Notes* for this release.
VLT LAG ID is not configured on one VLT peer	A syslog error message is generated. The peer with the VLT configured remains active.	A syslog error message is generated. The peer with the VLT configured remains active.	Verify the VLT LAG ID is configured correctly on both VLT peers.
VLT LAG ID mismatch	The VLT port channel is brought down. A syslog error message is generated.	The VLT port channel is brought down. A syslog error message is generated.	Perform a mismatch check after the VLT peer is established.

Reconfiguring Stacked Switches as VLT

To convert switches that have been stacked to VLT peers, use the following procedure.

1. Remove the current configuration from the switches. You will need to split the configuration up for each switch.
2. Copy the files to the flash memory of the appropriate switch.
3. Copy the files on the flash drive to the startup-config.
4. Reset the stacking ports to user ports for both switches.
5. Reload the stack and confirm the new configurations have been applied.
6. On the Secondary switch (stack-unit1), enter the command stack-unit1 renumber 0.
7. Confirm the reload query.
8. After reloading, confirm that VLT is enabled.
9. Confirm that the management ports are interconnected or connected to a switch that can transfer Heartbeat information.

Specifying VLT Nodes in a PVLAN

You can configure VLT peer nodes in a private VLAN (PVLAN). VLT enables redundancy without the implementation of Spanning Tree Protocol (STP), and provides a loop-free network with optimal bandwidth utilization.

Because the VLT LAG interfaces are terminated on two different nodes, PVLAN configuration of VLT VLANs and VLT LAGs are symmetrical and identical on both the VLT peers. PVLANs provide Layer 2 isolation between ports within the same VLAN. A PVLAN partitions a traditional VLAN into sub-domains identified by a primary and secondary VLAN pair. With VLT being a Layer 2 redundancy mechanism, support for configuration of VLT nodes in a PVLAN enables Layer 2 security functionalities. To achieve maximum VLT resiliency, you should configure the PVLAN IDs and mappings to be identical on both the VLT peer nodes.

The association of PVLAN with the VLT LAG must also be identical. After the VLT LAG is configured to be a member of either the primary or secondary PVLAN (which is associated with the primary), ICL becomes an automatic member of that PVLAN on both switches. This association helps the PVLAN data flow received on one VLT peer for a VLT LAG to be transmitted on that VLT LAG from the peer.

You can associate either a VLT VLAN or a VLT LAG to a PVLAN. First configure the VLT interconnect (VLTi) or a VLT LAG by using the `peer-link port-channel` *`id-number`* command or the VLT VLAN by using the `peer-link port-channel` *`id-number`* `peer-down-vlan vlan` *`interface number`* command and the `switchport` command. After you specify the VLTi link and VLT LAGs, you can associate the same port channel or LAG bundle that is a part of a VLT to a PVLAN by using the `interface` *`interface`* and `switchport mode private-vlan` commands.

When a VLTi port in trunk mode is a member of symmetric VLT PVLANs, the PVLAN packets are forwarded only if the PVLAN settings of both the VLT nodes are identical. You can configure the VLTi in trunk mode to be a member of non-VLT PVLANs if the VLTi is configured on both the peers. MAC address synchronization is performed for VLT PVLANs across peers in a VLT domain.

Keep the following points in mind when you configure VLT nodes in a PVLAN:

- Configure the VLTi link to be in trunk mode. Do not configure the VLTi link to be in access or promiscuous mode.
- You can configure a VLT LAG or port channel to be in trunk, access, or promiscuous port modes when you include the VLT LAG in a PVLAN. The VLT LAG settings must be the same on both the peers. If you configure a VLT LAG as a trunk port, you can associate that LAG to be a member of a normal VLAN or a PVLAN. If you configure a VLT LAG to be a promiscuous port, you can configure that LAG to be a member of PVLAN only. If you configure a VLT LAG to be in access port mode, you can add that LAG to be a member of the secondary VLAN only.
- ARP entries are synchronized even when a mismatch occurs in the PVLAN mode of a VLT LAG.

Any VLAN that contains at least one VLT port as a member is treated as a VLT VLAN. You can configure a VLT VLAN to be a primary, secondary, or a normal VLAN. However, the VLT VLAN configuration must be symmetrical across peers. If the VLT LAG is tagged to any one of the primary or secondary VLANs of a PVLAN, then both the primary and secondary VLANs are considered as VLT VLANs.

If you add an ICL or VLTi link as a member of a primary VLAN, the ICL becomes a part of the primary VLAN and its associated secondary VLANs, similar to the behavior for normal trunk ports. VLAN parity is not validated if you associate an ICL to a PVLAN. Similarly, if you dissociate an ICL from a PVLAN, although the PVLAN parity exists, ICL is removed from that PVLAN.

Association of VLTi as a Member of a PVLAN

If a VLAN is configured as a non-VLT VLAN on both the peers, the VLTi link is made a member of that VLAN if the VLTi link is configured as a PVLAN or normal VLAN on both the peers. If a PVLAN is configured as a VLT VLAN on one peer and a non-VLT VLAN on another peer, the VLTi is added as a member of that VLAN by verifying the PVLAN parity on both the peers. In such a case, if a PVLAN is present as a VLT PVLAN on at least one of the peers, then symmetric configuration of the PVLAN is validated to cause the VLTi to be a member of that VLAN. Whenever a change in the VLAN mode on one of the peers occurs, the information is synchronized with the other peer and VLTi is either added or removed from the VLAN based on the validation of the VLAN parity.

For VLT VLANs, the association between primary VLAN and secondary VLANs is examined on both the peers. Only if the association is identical on both the peers, VLTi is configured as a member of those VLANs. This behavior is because of security functionalities in a PVLAN. For example, if a VLAN is a primary VLT VLAN on one peer and not a primary VLT VLAN on the other peer, VLTi is not made a part of that VLAN.

MAC Synchronization for VLT Nodes in a PVLAN

For the MAC addresses that are learned on non-VLT ports, MAC address synchronization is performed with the other peer if the VLTi (ICL) link is part of the same VLAN as the non-VLT port. For MAC addresses that are learned on VLT ports, the VLT LAG mode of operation and the primary to secondary association of the VLT nodes is determined on both the VLT peers. MAC synchronization is performed for the VLT LAGs only if the VLT LAG and primary-secondary VLT peer mapping are symmetrical.

The PVLAN mode of VLT LAGs on one peer is validated against the PVLAN mode of VLT LAGs on the other peer. MAC addresses that are learned on that VLT LAG are synchronized between the peers only if the PVLAN mode on both the peers is identical. For example, if the MAC address is learned on a VLT LAG and the VLAN is a primary VLT VLAN on one peer and not a primary VLT VLAN on the other peer, MAC synchronization does not occur.

Whenever a change occurs in the VLAN mode of one of the peers, this modification is synchronized with the other peers. Depending on the validation mechanism that is initiated for MAC synchronization of VLT peers, MAC addresses learned on a particular VLAN are either synchronized with the other peers, or MAC addresses synchronized from the other peers on the same VLAN are deleted. This method of processing occurs when the PVLAN mode of VLT LAGs is modified.

Because the VLTi link is only a member of symmetric VLT PVLANs, MAC synchronization takes place directly based on the membership of the VLTi link in a VLAN and the VLT LAG mode.

PVLAN Operations When One VLT Peer is Down

When a VLT port moves to the Admin or Operationally Down state on only one of the VLT nodes, the VLT Lag is still considered to be up. All the PVLAN MAC entries that correspond to the operationally down VLT LAG are maintained as synchronized entries in the device. These MAC entries are removed when the peer VLT LAG also becomes inactive or a change in PVLAN configuration occurs.

PVLAN Operations When a VLT Peer is Restarted

When the VLT peer node is rebooted, the VLAN membership of the VLTi link is preserved and when the peer node comes back online, a verification is performed with the newly received PVLAN configuration from the peer. If any differences are identified, the VLTi link is either added or removed from the VLAN. When the peer node restarts and returns online, all the PVLAN configurations are exchanged across the peers. Based on the information received from the peer, a bulk synchronization of MAC addresses that belong to spanned PVLANs is performed.

During the booting phase or when the ICL link attempts to come up, a system logging message is recorded if VLT PVLAN mismatches, PVLAN mode mismatches, PVLAN association mismatches, or PVLAN port mode mismatches occur. Also, you can view these discrepancies if any occur by using the `show vlt mismatch` command.

Interoperation of VLT Nodes in a PVLAN with ARP Requests

When an ARP request is received, and the following conditions are applicable, the IP stack performs certain operations.

- The VLAN on which the ARP request is received is a secondary VLAN (community or isolated VLAN).
- Layer 3 communication between secondary VLANs in a private VLAN is enabled by using the `ip local-proxy-arp` command in INTERFACE VLAN configuration mode.
- The ARP request is not received on the ICL

Under such conditions, the IP stack performs the following operations:

- The ARP reply is sent with the MAC address of the primary VLAN.
- The ARP request packet originates on the primary VLAN for the intended destination IP address.

The ARP request received on ICLs are not proxied, even if they are received with a secondary VLAN tag. This behavior change occurs because the node from which the ARP request was forwarded would have replied with its MAC address, and the current node discards the ARP request.

Scenarios for VLAN Membership and MAC Synchronization With VLT Nodes in PVLAN

The following table illustrates the association of the VLTi link and PVLANs, and the MAC synchronization of VLT nodes in a PVLAN (for various modes of operations of the VLT peers):

Table 73. VLAN Membership and MAC Synchronization With VLT Nodes in PVLAN

VLT LAG Mode		PVLAN Mode of VLT VLAN		ICL VLAN Membership	Mac Synchronization
Peer1	Peer2	Peer1	Peer2		
Trunk	Trunk	Primary	Primary	Yes	Yes
Trunk	Trunk	Primary	Normal	No	No
Trunk	Trunk	Normal	Normal	Yes	Yes

VLT LAG Mode		PVLAN Mode of VLT VLAN		ICL VLAN Membership	Mac Synchronization
Peer1	Peer2	Peer1	Peer2		
Promiscuous	Trunk	Primary	Primary	Yes	No
Trunk	Access	Primary	Secondary	No	No
Promiscuous	Promiscuous	Primary	Primary	Yes	Yes
Promiscuous	Access	Primary	Secondary	No	No
Promiscuous	Promiscuous	Primary	Primary	Yes	Yes
		- Secondary (Community)	- Secondary (Isolated)	No	No
Access	Access	Secondary (Community)	Secondary (Isolated)	No	No
		• Primary X	• Primary X	Yes	Yes
Promiscuous	Promiscuous	Primary	Primary	Yes	Yes
		- Secondary (Community)	- Secondary (Community)	Yes	Yes
		- Secondary (Isolated)	- Secondary (Isolated)	Yes	Yes
Promiscuous	Trunk	Primary	Normal	No	No
Promiscuous	Trunk	Primary	Primary	Yes	No
Access	Access	Secondary (Community)	Secondary (Community)	Yes	Yes
		- Primary VLAN X	- Primary VLAN X	Yes	Yes
Access	Access	Secondary (Isolated)	Secondary (Isolated)	Yes	Yes
		- Primary VLAN X	- Primary VLAN X	Yes	Yes
Access	Access	Secondary (Isolated)	Secondary (Isolated)	No	No
		- Primary VLAN X	- Primary VLAN Y	No	No

VLT LAG Mode		**PVLAN Mode of VLT VLAN**		**ICL VLAN Membership**	**Mac Synchronization**
Peer1	**Peer2**	**Peer1**	**Peer2**		
Access	Access	Secondary (Community)	Secondary (Community)	No	No
		- Primary VLAN Y	- Primary VLAN X	No	No
Promiscuous	Access	Primary	Secondary	No	No
Trunk	Access	Primary/Normal	Secondary	No	No

Configuring a VLT VLAN or LAG in a PVLAN

You can configure the VLT peers or nodes in a private VLAN (PVLAN). Because the VLT LAG interfaces are terminated on two different nodes, PVLAN configuration of VLT VLANs and VLT LAGs are symmetrical and identical on both the VLT peers. PVLANs provide Layer 2 isolation between ports within the same VLAN. A PVLAN partitions a traditional VLAN into subdomains identified by a primary and secondary VLAN pair. With VLT being a Layer 2 redundancy feature, support for configuration of VLT nodes in a PVLAN enables Layer 2 security functionalities to be achieved. This section contains the following topics that describe how to configure a VLT VLAN or a VLT LAG (VLTi link) and assign that VLT interface to a PVLAN.

Creating a VLT LAG or a VLT VLAN

1. Configure the port channel for the VLT interconnect on a VLT switch and enter interface configuration mode

   CONFIGURATION mode

   ```
   interface port-channel id-number.
   ```

   Enter the same port-channel number configured with the `peer-link port-channel` command as described in Enabling VLT and Creating a VLT Domain.

   NOTE: To be included in the VLTi, the port channel must be in Default mode (`no switchport` or VLAN assigned).
2. Remove an IP address from the interface.

   INTERFACE PORT-CHANNEL mode

   ```
   no ip address
   ```
3. Add one or more port interfaces to the port channel.

   INTERFACE PORT-CHANNEL mode

   ```
   channel-member interface
   ```

   `interface`: specify one of the following interface types:

   - 1-Gigabit Ethernet: Enter `gigabitethernet` *`slot/port`*.
   - 10-Gigabit Ethernet: Enter `tengigabitethernet` *`slot/port`*.
4. Ensure that the port channel is active.

INTERFACE PORT-CHANNEL mode

```
no shutdown
```

5. To configure the VLT interconnect, repeat Steps 1–4 on the VLT peer switch.
6. Enter VLT-domain configuration mode for a specified VLT domain.

   CONFIGURATION mode

   ```
   vlt domain domain-id
   ```

   The range of domain IDs is from 1 to 1000.
7. Enter the port-channel number that acts as the interconnect trunk.

   VLT DOMAIN CONFIGURATION mode

   ```
   peer-link port-channel id-number
   ```

   The range is from 1 to 128.
8. (Optional) To configure a VLT LAG, enter the VLAN ID number of the VLAN where the VLT forwards packets received on the VLTi from an adjacent peer that is down.

   VLT DOMAIN CONFIGURATION mode

   ```
   peer-link port-channel id-number peer-down-vlan vlan interface number
   ```

   The range is from 1 to 4094.

Associating the VLT LAG or VLT VLAN in a PVLAN

1. Access INTERFACE mode for the port that you want to assign to a PVLAN.

   CONFIGURATION mode

   ```
   interface interface
   ```
2. Enable the port.

   INTERFACE mode

   ```
   no shutdown
   ```
3. Set the port in Layer 2 mode.

   INTERFACE mode

   ```
   switchport
   ```
4. Select the PVLAN mode.

   INTERFACE mode

   ```
   switchport mode private-vlan {host | promiscuous | trunk}
   ```

   - `host` (isolated or community VLAN port)
   - `promiscuous` (intra-VLAN communication port)
   - `trunk` (inter-switch PVLAN hub port)
5. Access INTERFACE VLAN mode for the VLAN to which you want to assign the PVLAN interfaces.

   CONFIGURATION mode

   ```
   interface vlan vlan-id
   ```

6. Enable the VLAN.

   INTERFACE VLAN mode

   ```
   no shutdown
   ```

7. To obtain maximum VLT resiliency, configure the PVLAN IDs and mappings to be identical on both the VLT peer nodes. Set the PVLAN mode of the selected VLAN to primary.

   INTERFACE VLAN mode

   ```
   private-vlan mode primary
   ```

8. Map secondary VLANs to the selected primary VLAN.

   INTERFACE VLAN mode

   ```
   private-vlan mapping secondary-vlan vlan-list
   ```

   The list of secondary VLANs can be:

   - Specified in comma-delimited (*`VLAN-ID,VLAN-ID`*) or hyphenated-range format (*`VLAN-ID-VLAN-ID`*).
   - Specified with this command even before they have been created.
   - Amended by specifying the new secondary VLAN to be added to the list.

Proxy ARP Capability on VLT Peer Nodes

The proxy ARP functionality is supported on VLT peer nodes.

A proxy ARP-enabled device answers the ARP requests that are destined for another host or router. The local host forwards the traffic to the proxy ARP-enabled device, which in turn transmits the packets to the destination.

By default, proxy ARP is enabled. To disable proxy ARP, use the `no proxy-arp` command in the interface mode. To re-enable proxy ARP, use the `ip proxy-arp` command in INTERFACE mode. To view if proxy ARP is enabled on the interface, use the `show config` command in INTERFACE mode. If it is not listed in the `show config` command output, it is enabled. Only nondefault information is displayed in the `show config` command output.

ARP proxy operation is performed on the VLT peer node IP address when the peer VLT node is down. The ARP proxy stops working either when the peer routing timer expires or when the peer VLT node goes up. Layer 3 VLT provides a higher resiliency at the Layer 3 forwarding level. VLT peer routing enables you to replace VRRP with routed VLT to route the traffic from Layer 2 access nodes. With proxy ARP, hosts can resolve the MAC address of the VLT node even when VLT node is down.

If the ICL link is down when a VLT node receives an ARP request for the IP address of the VLT peer, owing to LAG-level hashing algorithm in the top-of-rack (TOR) switch, the incorrect VLT node responds to the ARP request with the peer MAC address. Proxy ARP is not performed when the ICL link is up and the ARP request the wrong VLT peer. In this case, ARP requests are tunneled to the VLT peer.

Proxy ARP supported on both VLT interfaces and non-VLT interfaces. Proxy ARP supported on symmetric VLANs only. Proxy ARP is enabled by default. Routing table must be symmetrically configured to support proxy ARP. For example, consider a sample topology in which VLAN 100 is configured on two VLT nodes, node 1 and node 2. ICL link is not configured between the two VLT nodes. Assume that the VLAN 100 IP address in node 1 is 10.1.1.1/24 and VLAN 100 IP address in node 2 is 20.1.1.2/24. In this case, if the ARP request for 20.1.1.1 reaches node 1, node 1 will not perform the ARP request for 20.1.1.2. Proxy ARP is supported only for the IP address belongs to the received interface IP network. Proxy ARP is not supported if the ARP requested IP address is different from the received interface IP subnet. For example, if VLAN 100 and 200 are configured on the VLT peers, and if the VLAN 100 IP address is configured as 10.1.1.0/24 and the VLAN 200 IP address is configured as 20.1.1.0/24, the proxy ARP is not performed if the VLT node receives an ARP request for 20.1.1.0/24 on VLAN 100.

Working of Proxy ARP for VLT Peer Nodes

Proxy ARP is enabled only when peer routing is enabled on both the VLT peers. If peer routing is disabled on one of the VLT peers, proxy ARP is not performed when the ICL link goes down. Proxy ARP is performed only when the VLT peer's MAC address is installed in the database. Proxy ARP is stopped when the VLT peer's MAC address is removed from the ARP database because of the peer routing timer expiry. The source hardware address in the ARP response contains the VLT peer MAC address. Proxy ARP is supported for both unicast and broadcast ARP requests. Control packets, other than ARP requests destined for the VLT peers that reach the undesired and incorrect VLT node, are dropped if the ICL link is down. Further processing is not done on these control packets. The VLT node does not perform any action if it receives gratuitous ARP requests for the VLT peer IP address. Proxy ARP is also supported on secondary VLANs. When the ICL link or peer is down, and the ARP request for a private VLAN IP address reaches the wrong peer, then the wrong peer responds to the ARP request with the peer MAC address.

The IP address of the VLT node VLAN interface is synchronized with the VLT peer over ICL when the VLT peers are up. Whenever an IP address is added or deleted, this updated information is synchronized with the VLT peer. IP address synchronization occurs regardless of the VLAN administrative state. IP address addition and deletion serve as the trigger events for synchronization. When a VLAN state is down, the VLT peer might perform a proxy ARP operation for the IP addresses of that VLAN interface.

VLT nodes start performing Proxy ARP when the ICL link goes down. When the VLT peer comes up, proxy ARP will be stopped for the peer VLT IP addresses. When the peer node is rebooted, the IP address synchronized with the peer is not flushed. Peer down events cause the proxy ARP to commence.

When a VLT node detects peer up, it will not perform proxy ARP for the peer IP addresses. IP address synchronization occurs again between the VLT peers.

Proxy ARP is enabled only if peer routing is enabled on both the VLT peers. If you disable peer routing by using the `no peer-routing`command in VLT DOMAIN node, a notification is sent to the VLT peer to disable the proxy ARP. If peer routing is disabled when ICL link is down, a notification is not sent to the VLT peer and in such a case, the VLT peer does not disable the proxy ARP operation.

When the VLT domain is removed on one of the VLT nodes, the peer routing configuration removal will be notified to the peer. In this case VLT peer node disables the proxy ARP. When the ICL link is removed on one of the VLT nodes by using the `no peer-link` command, the ICL down event is triggered on the other VLT node, which in turn starts the proxy ARP application. The VLT node, where the ICL link is deleted, flushes the peer IP addresses and does not perform proxy ARP for the additional LAG hashed ARP requests.

VLT Nodes as Rendezvous Points for Multicast Resiliency

You can configure virtual link trunking (VLT) peer nodes as rendezvous points (RPs) in a Protocol Independent Multicast (PIM) domain.

PIM uses a VLT node as the RP to distribute multicast traffic to a multicast group. Messages to join the multicast group (Join messages) and data are sent towards the RP, so that receivers can discover who the senders are and begin receiving traffic destined for the multicast group.

To enable an explicit multicast routing table synchronization method for VLT nodes, you can configure VLT nodes as RPs. Multicast routing needs to identify the incoming interface for each route. The PIM running on both VLT peers enables both the peers to obtain traffic from the same incoming interface.

You can configure a VLT node to be an RP through the `ip pim rp-address` command in Global Configuration mode. When you configure a VLT node as an RP, the (*, G) routes that are synchronized from the VLT peers are ignored and not downloaded to the device. For the (S, G) routes that are synchronized from the VLT peer, after the RP starts receiving multicast traffic via these routes, these (S, G) routes are considered valid and are downloaded to the device. Only (S, G) routes are used to forward the multicast traffic from the source to the receiver.

You can configure VLT nodes, which function as RP, as Multicast Source Discovery Protocol (MSDP) peers in different domains. However, you cannot configure the VLT peers as MSDP peers in the same VLT domain. In such instances, the VLT peer does not support the RP functionality.

If the same source or RP can be accessed over both a VLT and a non-VLT VLAN, configure better metrics for the VLT VLANs. Otherwise, it is possible that one VLT node chooses a non-VLT VLAN (if the path through the VLT VLAN was not available when the route was learned) and another VLT node selects a VLT VLAN. Such a scenario can cause duplication of packets. ECMP is not supported when you configure VLT nodes as RPs.

Backup RP is not supported if the VLT peer that functions as the RP is statically configured. With static RP configuration, if the RP reboots, it can handle new clients only after it comes back online. Until the RP returns to the active state, the VLT peer forwards the packets for the already logged-in clients. To enable the VLT peer node to retain the synchronized multicast routes or synchronized multicast outgoing interface (OIF) maps after a peer node failure, use the timeout value that you configured through the `multicast peer-routing timeout` *`value`* command. You can configure an optimal time for a VLT node to retain synced multicast routes or synced multicast outgoing interface (OIF), after a VLT peer node failure, through the `multicast peer-routing-timeout` command in VLT DOMAIN mode. Using the bootstrap router (BSR) mechanism, both the VLT nodes in a VLT domain can be configured as the candidate RP for the same group range. When an RP fails, the VLT peer automatically takes over the role of the RP. This phenomenon enables resiliency to be achieved by the PIM BSR protocol.

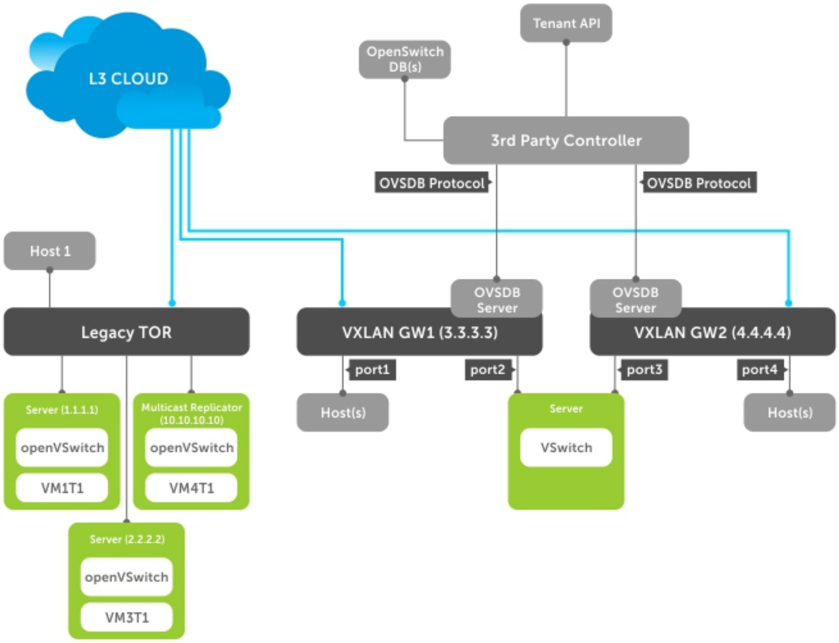
53

Virtual Extensible LAN (VXLAN)

Virtual Extensible LAN (VXLAN) is supported on the S6000 platform.

Overview

S6000 device acts as the VXLAN gateway and performs the VXLAN Tunnel End Point (VTEP) functionality. VXLAN is a technology where in the data traffic from the virtualized servers is transparently transported over an existing legacy network.


L3 CLOUD
OpenSwitch DB(s)
Tenant API
3rd Party Controller
OVSDB Protocol
OVSDB Protocol
Host 1
OVSDB Server
OVSDB Server
Legacy TOR
VXLAN GW1 (3.3.3.3)
VXLAN GW2 (4.4.4.4)
port1
port2
port3
port4
Server (1.1.1.1)
openVSwitch
VM1T1
Multicast Replicator (10.10.10.10)
openVSwitch
VM4T1
Host(s)
Server
VSwitch
Host(s)
Server (2.2.2.2)
openVSwitch
VM3T1

Components of VXLAN network

VXLAN provides a mechanism to extend an L2 network over an L3 network. In short, VXLAN is an L2 overlay scheme over an L3 network and this overlay is termed as a VXLAN segment.

Components of VXLAN network

The VXLAN network consists of the following components:

- Network Virtualization Platform (NVP) Controller
- VTEP (VXLAN Tunnel End Point)
- VXLAN Gateway
- VXLAN Hypervisor
- Service Node (SN)
- Legacy TOR

Network Virtualization Platform (NVP) Controller

NVP Controller is the network controller for managing cloud components. The OVSDB protocol is the protocol used for communication between VTEPs and the controller. In the current release, the qualified controller for the VXLAN Gateway function is NSX-from VMWare. The top-level functions of NVP are:

- Provide a GUI for creating service gateways.
- Manage the VTEPs.:
  - Binds Port and VLAN
  - Install VTEP tunnels
- Distribute the VTEPs to MAC binding to all relevant VTEPs.
- Provide an interface for cloud orchestration in cloud data center management.

VTEP (VXLAN Tunnel End Point)

VTEPs work as the open vSwitch running on the hypervisor on a virtualized server or as the VXLAN Gateway or as the Service Node (SN) that is responsible for flooding. The VTEPs are responsible for encapsulation and decapsulation of VXLAN headers.

VXLAN Gateways

VXLAN Gateways act as the VTEPs that encapsulate and decapsulate VXLAN headers. The roles and responsibilities of the Gateway are:

- Connects to the NVP client based on user configuration.
- Advertises south-facing VXLAN capable ports to the NVP client.
- Creates logical networks based on messages from the NVP.
- Creates tunnels to VTEPs based on messages from the NVP.
- Binds the Port and VLAN to logical networks based on messages from the NVP.
- Binds MACs to the VTEP and logical network based on messages from the NVP.
- Advertises MACs learnt on south-facing VXLAN capable-ports to the NVP client.

VXLAN Hypervisor

It is the VTEP that connects the Virtual Machines (VM) to the underlay legacy network to the physical infrastructure.

Service Node(SN)

It is also another VTEP, but it is fully managed by NSX. The purpose of SN is to be the central replication engine for flooded packets

Legacy TOR

It is a TOR switch, which performs routing or switching decisions.

Functional Overview of VXLAN Gateway

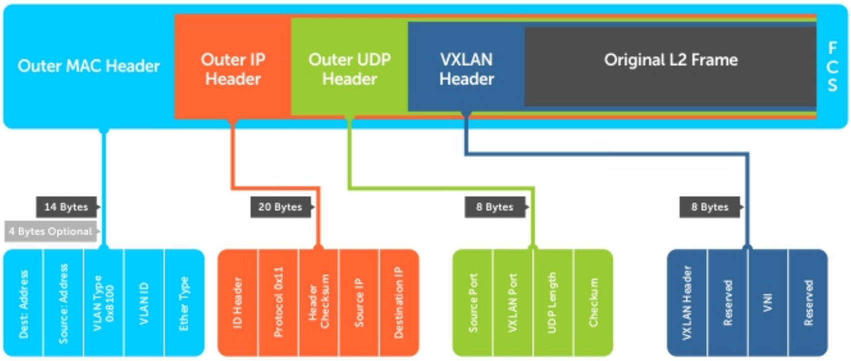
The following section is the functional overview of VXLAN Gateway:

1. Provides connectivity between a Virtual server infrastructure and a Physical server infrastructure.
2. Provides the functions performed by a VTEP in a virtual server infrastructure. The functions of a VTEP are:
    - VTEP is responsible for creating one or more logical networks.
    - VTEP is responsible for identifying and binding a Port and VLAN to a logical network
    - VTEP maintains MAC bindings to a VTEP.
    - VTEP is typically managed by a network orchestrator. In the case of S6000 being the VTEP, the VXLAN from VMWare is the network orchestrator.
    - VXLAN communicates with the S6000 VTEP via a standard protocol called OvsDb Protocol. The protocol uses the JSON RPC-based message format.
    - The S6000 VTEP acts according to the TOR schema defined by VMWare. The solution is very specific to VMWare-based orchestration platforms and does not work with other orchestration platforms.

VXLAN Frame Format

VXLAN provides a mechanism to extend an L2 network over an L3 network. In short, VXLAN is an L2 overlay scheme over an L3 network and this overlay is termed as a VXLAN segment. VMs within the same VXLAN segment communicate with each other through a 24–bit ID called VXLAN Network Identifier (VNI). This 24–bit ID allows up to 16 million VXLAN segments to coexist within the same administrative domain. The VNI scopes the inner MAC frame originated by the VM. This model has over-lapping MAC addresses without traffic cross-over

VXLAN achieves L2 over L3 overlay using the tunneling mechanism. VTEPs are the end points of a tunnel located within the hypervisor on the server that hosts VMs. The VXLAN frame format is shown in the following figure:


Outer MAC Header
Outer IP Header
Outer UDP Header
VXLAN Header
Original L2 Frame
FCS
14 Bytes
4 Bytes Optional
20 Bytes
8 Bytes
8 Bytes
Dest. Address
Source: Address
VLAN Type 0x8100
VLAN ID
Ether Type
ID Header
Protocol 0x11
Header Checksum
Source IP
Destination IP
Source Port
VXLAN Port
UDP Length
Checkum
VXLAN Header
Reserved
VNI
Reserved

Components of VXLAN Frame Format

Some of the important fields of the VXLAN frame format are described below:

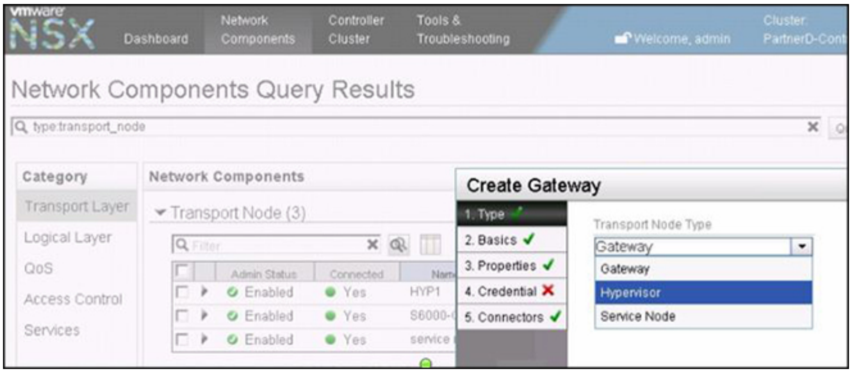
Outer Ethernet Header:	The Outer Ethernet Header consists of the following components: • **Destination Address**: Generally, it is a first hop router's MAC address when the VTEP is on a different address. • **Source Address** : It is the source MAC address of the router that routes the packet. • **VLAN**: It is optional in a VXLAN implementation and will be designated by an ethertype of 0×8100 and has an associated VLAN ID tag. • **Ethertype:** It is set to 0×0800 because the payload packet is an IPv4 packet. The initial VXLAN draft does not include an IPv6 implementation, but it is planned for the next draft.
Outer IP Header:	The Outer IP Header consists of the following components: • **Protocol**: It is set to 0×11 to indicate that the frame contains a UDP packet . • **Source IP**: It is the IP address of originating VTEP. • **Destination IP:** : It is the IP address of target VTEP.
Outer UDP Header:	The Outer UDP Header consists of the following components: • **Source Port:** Entropy of the inner frame. The entropy could be based on the Inner L2 header or Inner L3 header. • **VXLAN Port :** IANA-assigned VXLAN Port (4789). • **UDP Checksum:** The UDP checksum field is transmitted as zero. When a packet is received with a UDP checksum of zero, it is accepted for decapsulation.
VXLAN Header :	• **VXLAN Flags** : Reserved bits set to zero except bit 3, the first bit, which is set to 1 for a valid VNI • **VNI:** The 24-bit field that is the VXLAN Network Identifier • **Reserved:** A set of fields, 24 bits and 8 bits, that are reserved and set to zero .
Frame Check Sequence (FCS):	Note that the original Ethernet frame's FCS is not included, but new FCS is generated on the outer Ethernet frame.

Configuring and Controlling VXLAN from the NVP Controller GUI

To configure and control VXLAN from the NVP controller GUI, follow these steps:

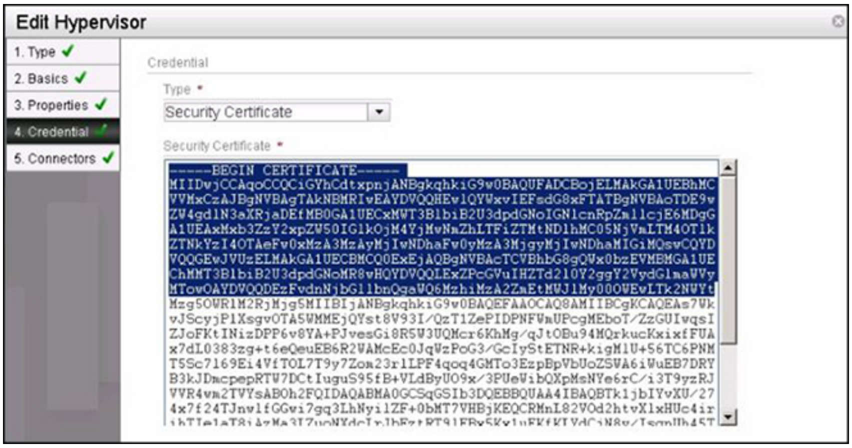
1. Create Hypervisor

To create a Hypervisor or server, the required fields are the IP address and SSL certificate of the server. The following are the snapshots of the user interface for creating a Hypervisor:


vmware NSX
Dashboard
Network Components
Controller Cluster
Tools & Troubleshooting
Welcome, admin
Cluster: PartnerD-Cont
Network Components Query Results
Category
Transport Layer
Logical Layer
QoS
Access Control
Services
Network Components
Transport Node (3)
Filter
Admin Status
Connected
Enabled
Yes
HYP1
S6000-
service
Create Gateway
1. Type
2. Basics
3. Properties
4. Credential
5. Connectors
Transport Node Type
Gateway
Hypervisor
Service Node

type:transport_node

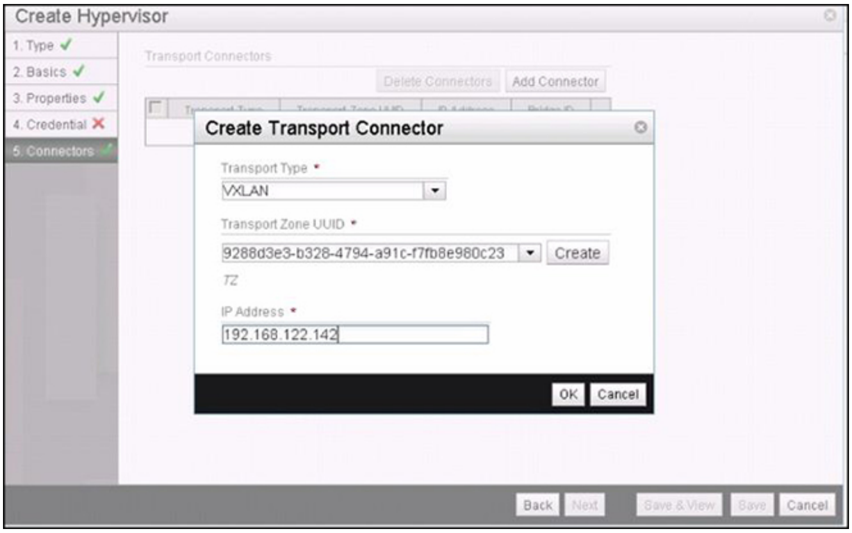
Transport Node Type: Gateway


Edit Hypervisor
1. Type
2. Basics
3. Properties
4. Credential
5. Connectors
Credential

Type *: Security Certificate

Security Certificate *:

```
-----BEGIN CERTIFICATE-----
MIIDvjCCAqoCCQCiGYhCdtxpnjANBgkqhkiG9w0BAQUFADCBojELMAkGA1UEBhMC
VVMxCzAJBgNVBAgTAkNBMRIwEAYDVQQHEwlQYWxvIEFsdG8xFTATBgNVBAoTDE9w
ZW4gdlN3aXRjaDEfMB0GA1UECxMWT3BlbiB2U3dpdGNoIGNlcnRpZmljYXRlcjE6MDgG
A1UEAxMxb3ZzY2xpZW50IGlkOjM4YjMvNmZhLTFiZTMtNDlhMC05NjVmLTM4OTlk
ZTNkYzI4OTAeFw0xMzA3MzAyMjIwNDhaFw0yMzA3MjgyMjIwNDhaMIGiMQswCQYD
VQQGEwJVUzELMAkGA1UECBMCQ0ExEjAQBgNVBAcTCVBhbG8gQWx0bzEVMBMGA1UE
ChMMT3BlbiB2U3dpdGNoMR8wHQYDVQQLExZPcGVuIHZTd2l0Y2ggY2VydGlmaWVy
MTowOAYDVQQDEzFvdnNjbGllbnQgaWQ6MzhiMzA2ZmEtMWJlMy00OWEwLTk2NWYt
Mzg5OWRlM2RjMjg5MIIBIjANBgkqhkiG9w0BAQEFAAOCAQ8AMIIBCgKCAQEAs7Wk
vJScyjPlXsgvOTA5WMMEjQYst8V93I/QzT1ZePIDPNFWmUPcgMEboT/ZzGUIvqsI
ZJoFKtINizDPP6v8YA+PJvesGi8R5W3UQMcr6KhMg/qJtOBu94MQrkucKxixfFUA
x7dL0383zg+t6eQeuEB6R2WAMcEc0JqWzPoG3/GcIyStETNR+kigMlU+56TC6PNM
T5Sc7169Ei4VfTOL7T9y7Zom23rlLPF4qoq4GMTo3EzpBpVbUoZSWA6iWuEB7DRY
B3kJDmcpepRTW7DCtIuguS95fB+VLdByUO9x/3PUeWibQXpMsNYe6rC/i3T9yzRJ
VVR4vm2TVYsABOh2FQIDAQABMA0GCSqGSIb3DQEBBQUAA4IBAQBTk1jbIYvXU/27
4x7f24TJnvlfGGwi7gq3LhNyilZF+0bMT7VHBjKEQCRMnL82VOd2htvXlxHUc4ir
```


Create Hypervisor
1. Type
2. Basics
3. Properties
4. Credential
5. Connectors
Transport Connectors
Delete Connectors
Add Connector
Create Transport Connector
Create
TZ
OK
Cancel
Back
Next
Save & View
Save
Cancel

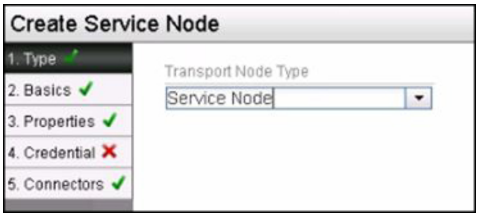
Transport Type *: VXLAN

Transport Zone UUID *: 9288d3e3-b328-4794-a91c-f7fb8e980c23

IP Address *: 192.168.122.142

2. Create Service Node

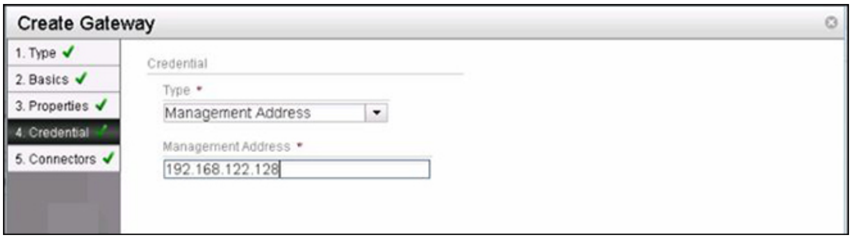
To create service node, the required fields are the IP address and SSL certificate of the server. The Service node is responsible for broadcast/unknown unicast/multicast traffic replication. The following is the snapshot of the user interface for the creation of service node:


Create Service Node
1. Type ✓
2. Basics ✓
3. Properties ✓
4. Credential ✗
5. Connectors ✓

Transport Node Type: Service Node

3. Create VXLAN Gateway

   To create a VXLAN L2 Gateway, the IP address of the Gateway is mandatory. The following is the snapshot of the user interface in creating a VXLAN Gateway

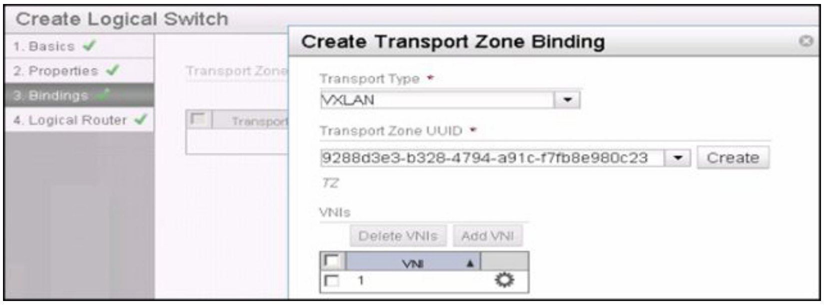

Create Gateway
1. Type ✓
2. Basics ✓
3. Properties ✓
4. Credential ✓
5. Connectors ✓
Credential

Type *: Management Address

Management Address *: 192.168.122.128

4. Create Logical Switch

   You can create a logical network by creating a logical switch. The logical network acts as the forwarding domain for workloads on the physical as well as virtual infrastructure.


Create Logical Switch
1. Basics ✓
2. Properties ✓
3. Bindings ✓
4. Logical Router ✓
Transport Zone
Transport
Create Transport Zone Binding
Create
Delete VNIs
Add VNI

Transport Type *: VXLAN

Transport Zone UUID *: 9288d3e3-b328-4794-a91c-f7fb8e980c23

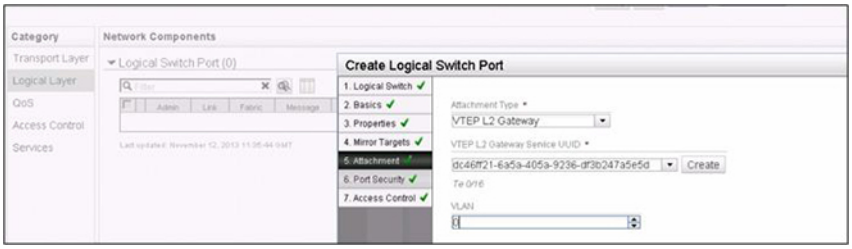
TZ

VNIs

☐	VNI
☐	1

5. Create Logical Switch Port

   A logical switch port provides a logical connection point for a VM interface (VIF) and a L2 gateway connection to an external network. It binds the virtual access ports in the GW to logical network (VXLAN) and VLAN.


Category
Transport Layer
Logical Layer
QoS
Access Control
Services
Network Components
Logical Switch Port (0)
Filter
Admin
Link
Fabric
Message
Last updated: November 12, 2013 11:26:44 GMT
Create Logical Switch Port
1. Logical Switch ✓
2. Basics ✓
3. Properties ✓
4. Mirror Targets ✓
5. Attachment ✓
6. Port Security ✓
7. Access Control ✓
Create

Attachment Type *: VTEP L2 Gateway

VTEP L2 Gateway Service UUID *: dc46ff21-6a5a-405a-9236-df3b247a5e5d

Te 0/16

VLAN: 0

NOTE: For more details about NVP controller configuration, refer to the NVP user guide from VMWare .

Configuring VxLAN Gateway

To configure the VxLAN gateway on the S6000 platform, follow these steps:

1. Connecting to NVP controller
2. Advertising VXLAN access ports to controller

Connecting to an NVP Controller

To connect to an NVP controller, use the following commands.

1. feature vxlan

   CONFIGURATION mode

   ```
   feature vxlan
   ```

   You must configure `feature VXLAN` to configure `vxlan-instance`.

2. vxlan-instance

   CONFIGURATION mode

   ```
   vxlan-instance instance ID
   ```

   The platform supports only the instance ID *1* in the initial release.

3. controller

   VxLAN INSTANCE mode

   controller *controller IDip address* port *port-number* tcp|ptcp|pssl|ssl

   The port number range is from 1 to 6632. The default port number is 6632. The default connection type is ssl.

4. gateway-ip

   VxLAN INSTANCE mode

   gateway-ip *IP address*

5. max-back off (Optional)

   VxLAN INSTANCE mode

   max_backoff *time*

   The range is from 1000-180000. The default value is 30000 milliseconds.

6. fail-mode (Optional)

   VxLAN INSTANCE mode

   fail-mode secure

   If the local VTEP loses connectivity with the controller, it will delete all its database and hardware flows/resources.
7. no shut

   VxLAN INSTANCE mode

Advertising VXLAN Access Ports to Controller

To advertise the access ports to the controller, use the following command.

In INTERFACE mode, `vxlan-instance` command configures a VXLAN-Access Port into a VXLAN-instance.

INTERFACE mode

`vxlan-instance`

Examples of the `show vxlan-instance` Command

```
Dell#show vxlan vxlan-instance 1
Instance        : 1
Admin State     : enabled
Management IP   : 192.168.200.200
Gateway IP      : 3.3.3.3
MAX Backoff     : 30000
Controller 1    : 192.168.122.6:6632 ssl (connected)
Fail Mode       : secure
Port List       :
  Fo 0/4    Te 0/16   Te 0/80   Po 2
```

The following example shows the `show vxlan vxlan-instance logical-network` command.

```
Dell#show vxlan vxlan-instance 1 logical-network
Instance        : 1
Total LN count  : 1
  Name                                                    VNID
  bffc3be0-13e6-4745-9f6b-0bcbc5877f01    4656

Dell#$n-instance 1 logical-network n 2a8d5d19-8845-4365-ad04-243f0b6df252
Name : 2a8d5d19-8845-4365-ad04-243f0b6df252
Description :
Tunnel Key : 2
VFI : 28674
Unknown Multicast MAC Tunnels:
    192.168.122.133 : vxlan_over_ipv4 (up)
Port Vlan Bindings:
    Te 0/80: VLAN: 0 (0x80000001),
    Fo 0/124: VLAN: 0 (0x80000004),
```

The following example shows the `show vxlan vxlan-instance statistics interface` command.

```
Dell#show vxlan vxlan-instance 1 statistics interface fortyGigE 0/124 100

Port : Fo 0/124 Vlan : 100

Rx Packets : 13
Rx Bytes : 1317

Tx Packets : 13

Tx Bytes : 1321
```

The following example shows the `show vxlan vxlan-instance physical-locator` command.

```
Dell#show vxlan vxlan-instance 1 physical-locator
Instance : 1
Tunnel   : count 1
    36.1.1.1 : vxlan_over_ipv4 (up)
```

The following example shows the `show vxlan vxlan-instance unicast-mac-local` command.

```
Dell# show vxlan vxlan-instance <1> unicast-mac-local
Total Local Mac Count:    5
VNI              MAC                          PORT      VLAN

4656            00:00:02:00:03:00     Te 0/17   0
4656            00:00:02:00:03:01     Te 0/17   0
4656            00:00:02:00:03:02     Te 0/17   0
4656            00:00:02:00:03:03     Te 0/17   0
4656            00:00:02:00:03:04     Te 0/17   0
```

The following example shows the `show vxlan vxlan-instance unicast-mac-remote` command.

```
Dell# show vxlan vxlan-instance <1> unicast-mac-remote
Total Local Mac Count:    1
VNI             MAC                          TUNNEL
4656           00:00:01:00:00:01     36.1.1.1
```

Displaying VXLAN Configurations

To display the VXLAN configurations, use the following commands.

Examples of the `show vxlan-instance` Command

The following example shows the `show vxlan vxlan-instance` command.

```
Dell#show vxlan vxlan-instance 1
Instance        : 1
Admin State     : enabled
Management IP   : 192.168.200.200
Gateway IP      : 3.3.3.3
MAX Backoff     : 30000
Controller 1    : 192.168.122.6:6632 ssl (connected)
Fail Mode       : secure
Port List       :
  Fo 0/4    Te 0/16   Te 0/80   Po 2
```

The following example shows the `show vxlan vxlan-instance logical-network` command.

```
Dell#show vxlan vxlan-instance 1 logical-network
Instance        : 1
Total LN count  : 1
  Name                                                     VNID
  bffc3be0-13e6-4745-9f6b-0bcbc5877f01    4656

Dell#$n-instance 1 logical-network n 2a8d5d19-8845-4365-ad04-243f0b6df252
Name : 2a8d5d19-8845-4365-ad04-243f0b6df252
Description :
```

```
Tunnel Key : 2
VFI : 28674
Unknown Multicast MAC Tunnels:
    192.168.122.133 : vxlan_over_ipv4 (up)
Port Vlan Bindings:
    Te 0/80: VLAN: 0 (0x80000001),
    Fo 0/124: VLAN: 0 (0x80000004),
```

The following example shows the `show vxlan vxlan-instance statistics interface` command.

```
Dell#show vxlan vxlan-instance 1 statistics interface fortyGigE 0/124 100

Port : Fo 0/124 Vlan : 100

Rx Packets : 13

Rx Bytes : 1317

Tx Packets : 13

Tx Bytes : 1321
```

The following example shows the `show vxlan vxlan-instance physical-locator` command.

```
Dell#show vxlan vxlan-instance 1 physical-locator
Instance : 1
Tunnel   : count 1
    36.1.1.1 : vxlan_over_ipv4 (up)
```

The following example shows the `show vxlan vxlan-instance unicast-mac-local` command.

```
Dell# show vxlan vxlan-instance <1> unicast-mac-local
Total Local Mac Count:    5
VNI              MAC                        PORT      VLAN

4656            00:00:02:00:03:00     Te 0/17   0
4656            00:00:02:00:03:01     Te 0/17   0
4656            00:00:02:00:03:02     Te 0/17   0
4656            00:00:02:00:03:03     Te 0/17   0
4656            00:00:02:00:03:04     Te 0/17   0
```

The following example shows the `show vxlan vxlan-instance unicast-mac-remote` command.

```
Dell# show vxlan vxlan-instance <1> unicast-mac-remote
Total Local Mac Count:    1
VNI             MAC                        TUNNEL
4656           00:00:01:00:00:01     36.1.1.1
```

54

Virtual Routing and Forwarding (VRF)

Virtual Routing and Forwarding (VRF) allows a physical router to partition itself into multiple Virtual Routers (VRs). The control and data plane are isolated in each VR so that traffic does NOT flow across VRs.Virtual Routing and Forwarding (VRF) allows multiple instances of a routing table to co-exist within the same router at the same time.

VRF Overview

VRF improves functionality by allowing network paths to be segmented without using multiple devices. Using VRF also increases network security and can eliminate the need for encryption and authentication due to traffic segmentation.

Internet service providers (ISPs) often take advantage of VRF to create separate virtual private networks (VPNs) for customers; VRF is also referred to as VPN routing and forwarding.

VRF acts like a logical router; while a physical router may include many routing tables, a VRF instance uses only a single routing table. VRF uses a forwarding table that designates the next hop for each data packet, a list of devices that may be called upon to forward the packet, and a set of rules and routing protocols that govern how the packet is forwarded. These VRF forwarding tables prevent traffic from being forwarded outside a specific VRF path and also keep out traffic that should remain outside the VRF path.

VRF uses interfaces to distinguish routes for different VRF instances. Interfaces in a VRF can be either physical (Ethernet port or port channel) or logical (VLANs). You can configure identical or overlapping IP subnets on different interfaces if each interface belongs to a different VRF instance.

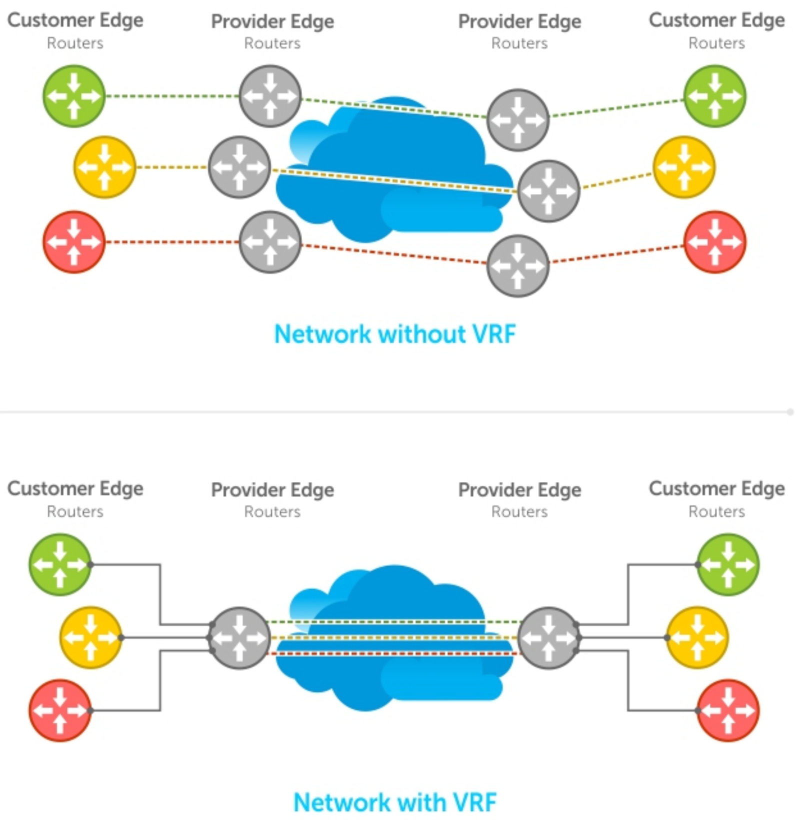

Customer Edge
Routers
Provider Edge
Routers
Provider Edge
Routers
Customer Edge
Routers
Network without VRF
Customer Edge
Routers
Provider Edge
Routers
Provider Edge
Routers
Customer Edge
Routers
Network with VRF

Figure 122. VRF Network Example

VRF Configuration Notes

Although there is no restriction on the number of VLANs that can be assigned to a VRF instance, the total number of routes supported in VRF is limited by the size of the IPv4 CAM.

VRF is implemented in a network device by using Forwarding Information Bases (FIBs).

A network device may have the ability to configure different virtual routers, where entries in the FIB that belong to one VRF cannot be accessed by another VRF on the same device. Only Layer 3 interfaces can belong to a VRF. VRF is supported on following types of interface:

- Physical Ethernet interfaces
- Port-channel interfaces (static & dynamic using LACP)
- VLAN interfaces
- Loopback interfaces

VRF supports route redistribution between routing protocols (including static routes) only when the routes are within the same VRF.

Dell Networking OS uses both the VRF name and VRF ID to manage VRF instances. The VRF name and VRF ID number are assigned using the ip vrf command. The VRF ID is displayed in show ip vrf command output.

The VRF ID is not exchanged between routers. VRF IDs are local to a router.

VRF supports some routing protocols only on the default VRF (default-vrf) instance. Table 1 displays the software features supported in VRF and whether they are supported on all VRF instances or only the default VRF.

Table 74.

Feature/Capability	Support Status for Default VRF	Support Status for Non-default VRF
Configuration rollback for commands introduced or modified	Yes	No
LLDP protocol on the port	Yes	No
802.1x protocol on the VLAN port	Yes	No
OSPF, RIP, ISIS, BGP on physical and logical interfaces	Yes	Yes NOTE: OSPF supported on all VRF ports. OSPF V2 and BGP V4 are supported on non-default-VRF ports also. Others supported only on default-VRF ports.
Dynamic Port-channel (LACP) on VLAN port or a Layer 3 port	Yes	Yes
Static Port-channel as VLAN port or a Layer 3 port	Yes	Yes
Port-monitoring	Yes	No
BFD on physical and logical interfaces	Yes	No
PVST, MSTP, RSTP and 802.1D STP for VLANs	Yes	No
FRRP (if applicable) for VLANs	Yes	No
Multicast protocols (PIM-SM, PIM-DM, MSDP)	Yes	No
Layer 3 (IPv4/IPv6) ACLs, TraceLists, PBR, QoS on VLANs	Yes	Yes

Feature/Capability	Support Status for Default VRF	Support Status for Non-default VRF
		NOTE: ACLs supported on all VRF VLAN ports. IPv4 ACLs are supported on non-default-VRFs also. IPv6 ACLs are supported on default-VRF only. PBR supported on default-VRF only. QoS not supported on VLANs.
Layer 3 (IPv4/IPv6) ACLs, TraceLists, PBR, QoS on physical interfaces and LAGs	Yes	Yes NOTE: ACLs supported on all VRF ports. TraceLists are common for entire line card (except on ExaScale). PBR supported on default-VRF only. QoS supported on all VRF ports.
IPv4 ARP	Yes	Yes
IPv6 Neighbor Discovery	Yes	No
Layer 2 ACLs on VLANs	Yes	No
FEED	Yes	No
Layer 2 QoS	Yes	Yes
Support for storm-control (broadcast and unknown-unicast)	Yes	No
sFlow	Yes	No
VRRP on physical and logical interfaces	Yes	Yes
Secondary IP Addresses	Yes	No
Following IPv6 capabilities		No
Basic	Yes	No
OSPFv3	Yes	No
ISIS	Yes	No
BGP	Yes	No
ACL	Yes	Yes
Multicast	Yes	No
NDP	Yes	No

Feature/Capability	Support Status for Default VRF	Support Status for Non-default VRF
RAD	Yes	No
Ingress/Egress Storm-Control (per-interface/global)	Yes	No

DHCP

DHCP requests are not forwarded across VRF instances. The DHCP client and server must be on the same VRF instance.

VRF Configuration

The VRF configuration tasks are:

1. Enabling VRF in Configuration Mode
2. Creating a Non-Default VRF
3. Assign an Interface to a VRF

You can also:

- View VRF Instance Information
- Connect an OSPF Process to a VRF Instance
- Configure VRRP on a VRF

Load VRF CAM

VRF is enabled by default on the switch. To load the VRF CAM profile, enter the `feature vrf` command in global configuration mode.

Step	Task	Command Syntax	Command Mode
1	Load CAM memory for the VRF feature.	`feature vrf`	CONFIGURATION

After you load VRF CAM, CLI parameters that allow you to configure non-default VRFs are made available on the system.

Creating a Non-Default VRF Instance

VRF is enabled by default on the switch and supports up to 64 VRF instances: 1 to 63 and the default VRF (0).

Task	Command Syntax	Command Mode
Create a non-default VRF instance by specifying a name and VRF ID number, and enter VRF configuration mode.	`ip vrf` *`vrf-name vrf-id`* VRF ID range: 1 to 63 and 0 (default VRF)	CONFIGURATION

Assigning an Interface to a VRF

You must enter the `ip vrf forwarding` command before you configure the IP address or any other setting on an interface.

NOTE: You can configure an IP address or subnet on a physical or VLAN interface that overlaps the same IP address or subnet configured on another interface only if the interfaces are assigned to different VRFs. If two interfaces are assigned to the same VRF, you cannot configure overlapping IP subnets or the same IP address on them.

Task	**Command Syntax**	**Command Mode**
Assign an interface to a VRF instance.	`ip vrf forwarding` *`vrf-name`*	INTERFACE

View VRF Instance Information

To display information about VRF configuration, enter the `show ip vrf` command.

Task	**Command Syntax**	**Command Mode**
Display the interfaces assigned to a VRF instance. To display information on all VRF instances (including the default VRF 0), do not enter a value for *`vrf-name`*.	`show ip vrf [`*`vrf-name`*`]`	EXEC

Assigning an OSPF Process to a VRF Instance

OSPF routes are supported on all VRF instances. Refer to Open Shortest Path First (OSPFv2) for complete OSPF configuration information.

Assign an OSPF process to a VRF instance . Return to CONFIGURATION mode to enable the OSPF process. The OSPF Process ID is the identifying number assigned to the OSPF process, and the Router ID is the IP address associated with the OSPF process.

Task	**Command Syntax**	**Command Mode**
Enable the OSPFv2 process globally for a VRF instance. Enter the VRF key word and instance name to tie the OSPF instance to the VRF. All network commands under this OSPF instance are subsequently tied to the VRF instance. *process-id range*: 0-65535	`router ospf` *`process-id`* `vrf` *`vrf name`*	CONFIGURATION

Once the OSPF process and the VRF are tied together, the OSPF Process ID cannot be used again in the system.

Configuring VRRP on a VRF Instance

You can configure the VRRP feature on interfaces that belong to a VRF instance.

In a virtualized network that consists of multiple VRFs, various overlay networks can exist on a shared physical infrastructure. Nodes (hosts and servers) that are part of the VRFs can be configured with IP static routes for reaching specific destinations through a given gateway in a VRF. VRRP provides high availability and protection for next-hop static routes by eliminating a single point of failure in the default static routed network. For more information, refer to VRRP Overview.

Task	Command Syntax	Command Mode
Create VRF	`ip vrf vrf1`	CONFIGURATION
Assign the VRF to an interface	`ip vrf forwarding vrf1`	VRF CONFIGURATION
Assign an IP address to the interface	`ip address 10.1.1.1 /24` `no shutdown`	
Configure the VRRP group and virtual IP address	see below	
View VRRP command output for the VRF vrf1	see below	

Configure the VRRP group and virtual IP address

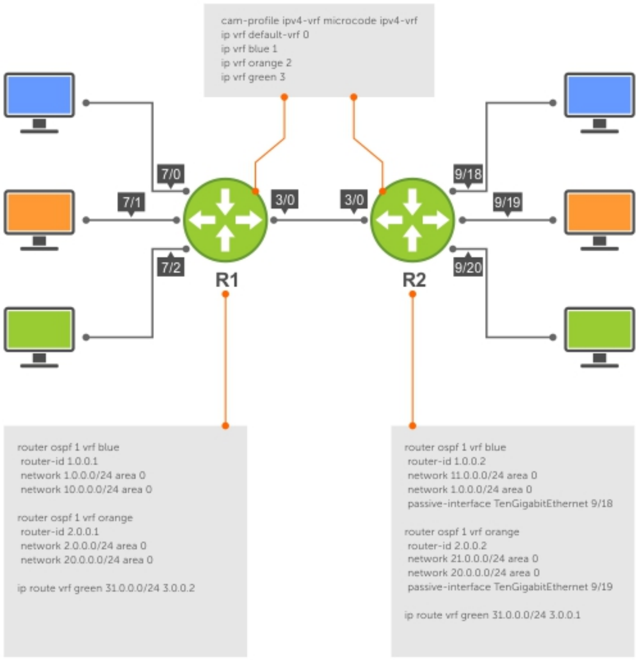
```
vrrp-group 10
virtual-address 10.1.1.100
show config
-----------------------------
!
interface TenGigabitEthernet 0/13
ip vrf forwarding vrf1
ip address 10.1.1.1/24
!
 vrrp-group 10
  virtual-address 10.1.1.100
 no shutdown
```

View VRRP command output for the VRF vrf1

```
show vrrp vrf vrf1
------------------
TenGigabitEthernet 0/13, IPv4
VRID: 10, Version: 2, Net:
10.1.1.1
VRF: 2 vrf1
State: Master, Priority: 100,
Master: 10.1.1.1 (local)
Hold Down: 0 sec, Preempt: TRUE,
AdvInt: 1 sec
Adv rcvd: 0, Bad pkts rcvd: 0,
Adv sent: 43, Gratuitous ARP
sent: 0
Virtual MAC address:
00:00:5e:00:01:0a
Virtual IP address:
10.1.1.100
Authentication: (none)
```

Sample VRF Configuration

The following configuration illustrates a typical VRF set-up.

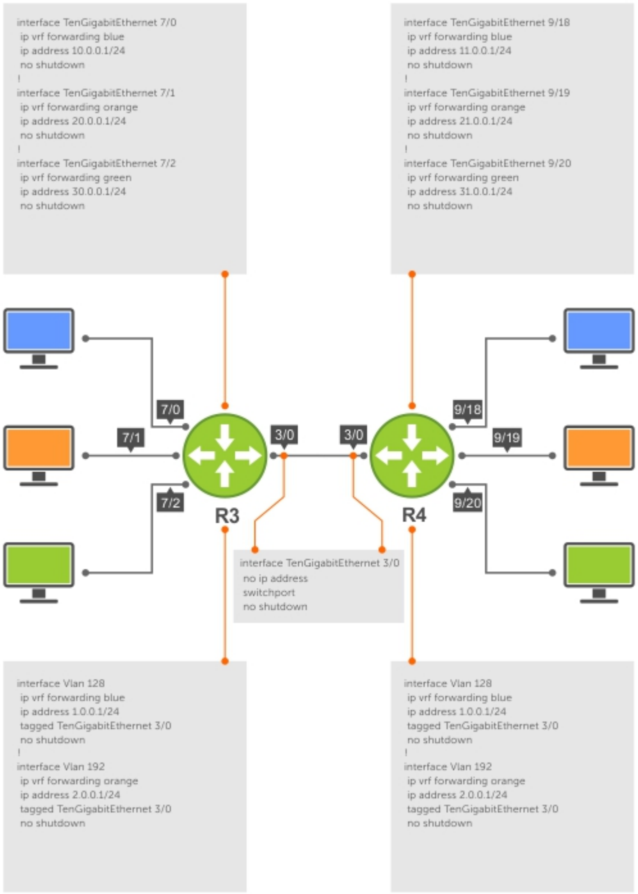


```
cam-profile ipv4-vrf microcode ipv4-vrf
ip vrf default-vrf 0
ip vrf blue 1
ip vrf orange 2
ip vrf green 3
```

```
router ospf 1 vrf blue
 router-id 1.0.0.1
 network 1.0.0.0/24 area 0
 network 10.0.0.0/24 area 0

router ospf 1 vrf orange
 router-id 2.0.0.1
 network 2.0.0.0/24 area 0
 network 20.0.0.0/24 area 0

ip route vrf green 31.0.0.0/24 3.0.0.2
```

```
router ospf 1 vrf blue
 router-id 1.0.0.2
 network 11.0.0.0/24 area 0
 network 1.0.0.0/24 area 0
 passive-interface TenGigabitEthernet 9/18

router ospf 1 vrf orange
 router-id 2.0.0.2
 network 21.0.0.0/24 area 0
 network 20.0.0.0/24 area 0
 passive-interface TenGigabitEthernet 9/19

ip route vrf green 31.0.0.0/24 3.0.0.1
```

Figure 123. Setup OSPF and Static Routes



```
interface TenGigabitEthernet 7/0
 ip vrf forwarding blue
 ip address 10.0.0.1/24
 no shutdown
!
interface TenGigabitEthernet 7/1
 ip vrf forwarding orange
 ip address 20.0.0.1/24
 no shutdown
!
interface TenGigabitEthernet 7/2
 ip vrf forwarding green
 ip address 30.0.0.1/24
 no shutdown
```

```
interface TenGigabitEthernet 9/18
 ip vrf forwarding blue
 ip address 11.0.0.1/24
 no shutdown
!
interface TenGigabitEthernet 9/19
 ip vrf forwarding orange
 ip address 21.0.0.1/24
 no shutdown
!
interface TenGigabitEthernet 9/20
 ip vrf forwarding green
 ip address 31.0.0.1/24
 no shutdown
```

```
interface TenGigabitEthernet 3/0
 no ip address
 switchport
 no shutdown
```

```
interface Vlan 128
 ip vrf forwarding blue
 ip address 1.0.0.1/24
 tagged TenGigabitEthernet 3/0
 no shutdown
!
interface Vlan 192
 ip vrf forwarding orange
 ip address 2.0.0.1/24
 tagged TenGigabitEthernet 3/0
 no shutdown
```

```
interface Vlan 128
 ip vrf forwarding blue
 ip address 1.0.0.1/24
 tagged TenGigabitEthernet 3/0
 no shutdown
!
interface Vlan 192
 ip vrf forwarding orange
 ip address 2.0.0.1/24
 tagged TenGigabitEthernet 3/0
 no shutdown
```

Figure 124. Setup VRF Interfaces

The following example relates to the configuration shown in Figure1 and Figure 2.

Router 1

```
ip vrf blue 1
!
ip vrf orange 2
!
ip vrf green 3
!
interface TenGigabitEthernet 3/0
    no ip address
    switchport
    no shutdown
!
interface GigabitEthernet 7/0
    ip vrf forwarding blue
    ip address 10.0.0.1/24
    no shutdown
!
```

```
interface GigabitEthernet 7/1
    ip vrf forwarding orange
    ip address 20.0.0.1/24
    no shutdown
!
interface GigabitEthernet 7/2
    ip vrf forwarding green
    ip address 30.0.0.1/24
    no shutdown
!
interface Vlan 128
    ip vrf forwarding blue
    ip address 1.0.0.1/24
    tagged TenGigabitEthernet 3/0
    no shutdown
!
interface Vlan 192
    ip vrf forwarding orange
    ip address 2.0.0.1/24
    tagged TenGigabitEthernet 3/0
    no shutdown
!
interface Vlan 256
    ip vrf forwarding green
    ip address 3.0.0.1/24
    tagged TenGigabitEthernet 3/0
    no shutdown
!
router ospf 1 vrf blue
    router-id 1.0.0.1
    network 1.0.0.0/24 area 0
    network 10.0.0.0/24 area 0
!
router ospf 2 vrf orange
    router-id 2.0.0.1
    network 2.0.0.0/24 area 0
    network 20.0.0.0/24 area 0
!
ip route vrf green 31.0.0.0/24 3.0.0.2!

================================================================================
================================
```

Router 2

```
ip vrf blue 1
!
ip vrf orange 2
!
ip vrf green 3
!
interface TenGigabitEthernet 3/0
    no ip address
    switchport
    no shutdown
!
interface GigabitEthernet 9/18
    ip vrf forwarding blue
    ip address 11.0.0.1/24
    no shutdown
!
```

```
interface GigabitEthernet 9/19
    ip vrf forwarding orange
    ip address 21.0.0.1/24
    no shutdown
!
interface GigabitEthernet 9/20
    ip vrf forwarding green
    ip address 31.0.0.1/24
    no shutdown
!
interface Vlan 128
    ip vrf forwarding blue
    ip address 1.0.0.2/24
    tagged TenGigabitEthernet 3/0
no shutdown
interface Vlan 192
    ip vrf forwarding orange
    ip address 2.0.0.2/24
    tagged TenGigabitEthernet 3/0
    no shutdown
!
interface Vlan 256
    ip vrf forwarding green
    ip address 3.0.0.2/24
    tagged TenGigabitEthernet 3/0
    no shutdown
!
router ospf 1 vrf blue
    router-id 1.0.0.2
    network 11.0.0.0/24 area 0
    network 1.0.0.0/24 area 0
    passive-interface GigabitEthernet 9/18
!
router ospf 2 vrf orange
    router-id 2.0.0.2
    network 21.0.0.0/24 area 0
    network 2.0.0.0/24 area 0
    passive-interface GigabitEthernet 9/19
!
ip route vrf green30.0.0.0/24 3.0.0.1
!

================================================================================
=================
```

The following shows the output of the show commands on Router 1.

Router 1

```
Dell#show ip vrf
VRF-Name                        VRF-ID Interfaces

default-vrf                     0       Gi 2/0-89,
                                        Te 3/0-3,
                                        Gi 4/0-89,
                                        Gi 5/0-89,
                                        Gi 7/3-47,
                                        Gi 9/0-47,
                                        Gi 10/0-47,
                                        Gi 11/0-47,
                                        Gi 12/0-47,
                                         Gi 13/0-47,
                                         Ma 0/0,
                                         Ma 1/0,
                                         Nu 0,
                                         Vl 1
blue                             1       Gi 7/0,
                                         Vl 128
orange                           2       Gi 7/1,
                                         Vl 192
green                            3       Gi 7/2,
                                         Vl 256
Dell#show ip ospf 1 neighbor
Neighbor ID     Pri     State         Dead Time Address Interface Area
1.0.0.2          1     FULL/DR     00:00:32 1.0.0.2         Vl 128         0

Dell#sh ip ospf 2 neighbor
Neighbor ID  Pri     State         Dead Time Address Interface Area
2.0.0.2          1     FULL/DR     00:00:37 2.0.0.2         Vl 192         0
Dell#show ip route vrf blue

Codes: C - connected, S - static, R - RIP,
            B - BGP, IN - internal BGP, EX - external BGP,LO - Locally
Originated,
            O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
            N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
            E2 - OSPF external type 2, i - IS-IS, L1 - IS-IS level-1,
            L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
            > - non-active route, + - summary route

Gateway of last resort is not set

        Destination         Gateway                 Dist/Metric         Last
Change
        -----------        -------                 -----------
-----------
C         1.0.0.0/24         Direct, Vl 128     0/0                     00:20:48
C         10.0.0.0/24         Direct, Gi 7/0     0/0
00:10:06
O         11.0.0.0/24         via 1.0.0.2,         Vl 128 110/2     00:11:13

Dell#show ip route vrf orange

Codes: C - connected, S - static, R - RIP,
            B - BGP, IN - internal BGP, EX - external BGP,LO - Locally
Originated,
            O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
            N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
            E2 - OSPF external type 2, i - IS-IS, L1 - IS-IS level-1,
            L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
            > - non-active route, + - summary route

Gateway of last resort is not set

        Destination         Gateway                         Dist/Metric
Last Change
        -----------         -------                         -----------
-----------
C         2.0.0.0/24         Direct, Vl 192             0/0
00:20:55
C         20.0.0.0/24         Direct, Gi 7/1
0/0                     00:10:05
O         21.0.0.0/24         via 2.0.0.2,    Vl 192
```

```
110/2               00:10:41

Dell#show ip route vrf green

Codes: C - connected, S - static, R - RIP,
            B - BGP, IN - internal BGP, EX - external BGP,LO - Locally
Originated,
            O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
            N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
            E2 - OSPF external type 2, i - IS-IS, L1 - IS-IS level-1,
            L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
            > - non-active route, + - summary route

Gateway of last resort is not set

        Destination         Gateway                        Dist/Metric
Last Change
        -----------         -------                        -----------
-----------
C         3.0.0.0/24         Direct, Vl 256            0/0
00:20:52
C         30.0.0.0/24         Direct, Gi 7/2
0/0                    00:09:45
S         31.0.0.0/24         via 3.0.0.2, Vl 256     1/0
00:09:06
================================================================================
======
```

The following shows the output of the show commands on Router 2.

Router 2

```
Dell#show ip vrf
VRF-Name                                                 VRF-ID     Interfaces
default-vrf                                             0             Gi 1/0-89,
                                                                        Te
3/0-3,
                                                                        Gi
4/0-89,
                                                                        Gi
5/0-89,
                                                                        Gi
6/0-89,
                                                                        Gi
9/0-17,21-47,
                                                                        Gi
11/0-47,
                                                                        Gi
12/0-47,
                                                                        Gi
13/0-47,
                                                                        Ma 0/0,
                                                                        Ma 1/0,
                                                                        Nu 0,
                                                                        Vl 1
blue                                                     1            Gi 9/18,
                                                                        Vl 128
orange                                                     2            Gi
9/19,
                                                                        Vl 192
green                                                     3             Gi 9/20,
                                                                      Vl 256

Dell#show ip ospf 1 neighbor
Neighbor ID         Pri             State Dead Time     Address
Interface     Area
1.0.0.1 1         FULL/BDR     00:00:36                1.0.0.1     Vl
128             0
!
Dell#sh ip ospf 2 neighbor
Neighbor ID         Pri             State Dead Time     Address
Interface     Area
2.0.0.1 1         FULL/BDR     00:00:33                2.0.0.1     Vl
192             0
!
Dell#show ip route vrf blue

Codes: C - connected, S - static, R - RIP,
            B - BGP, IN - internal BGP, EX - external BGP,LO - Locally
Originated,
            O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
            N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
            E2 - OSPF external type 2, i - IS-IS, L1 - IS-IS level-1,
            L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
            > - non-active route, + - summary route

Gateway of last resort is not set
        Destination         Gateway                         Dist/Metric
Last Change
        -----------         -------                         -----------
-----------
C         1.0.0.0/24         Direct,    Vl 128
0/0                     00:27:21
O         10.0.0.0/24         via 1.0.0.1,    Vl 128
110/2                 00:14:24
C         11.0.0.0/24         Direct, Gi 9/18
0/0                     00:19:46

Dell#show ip route vrf orange

Codes: C - connected, S - static, R - RIP,
            B - BGP, IN - internal BGP, EX - external BGP,LO - Locally
Originated,
            O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
            N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
            E2 - OSPF external type 2, i - IS-IS, L1 - IS-IS level-1,
            L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
            > - non-active route, + - summary route

Gateway of last resort is not set

        Destination         Gateway                         Dist/Metric
Last Change
        -----------         -------                         -----------
-----------
C         2.0.0.0/24         Direct, Vl 192             0/0
00:26:44
O         20.0.0.0/24         via 2.0.0.1, Vl 192     110/2
00:14:22
C         21.0.0.0/24         Direct, Gi 9/19
0/0                     00:20:38

Dell#show ip route vrf green
```

```
Codes: C - connected, S - static, R - RIP,
            B - BGP, IN - internal BGP, EX - external BGP,LO - Locally
Originated,
            O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
            N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
            E2 - OSPF external type 2, i - IS-IS, L1 - IS-IS level-1,
            L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
            > - non-active route, + - summary route

Gateway of last resort is not set

        Destination         Gateway                         Dist/Metric
Last Change
        -----------         -------                         -----------
-----------
C         3.0.0.0/24         Direct, Vl 256             0/0
00:26:27
S         30.0.0.0/24         via 3.0.0.1, Vl 256     1/0
00:17:03
C         31.0.0.0/24         Direct, Gi 9/20
0/0                     00:20:19
Dell#
```

Route Leaking VRFs

Static routes can be used to redistribute routes between non-default to default/non-default VRF and vice-versa.

You can configure route leaking between two VRFs using the following command: `ip route vrf x.x.x.x s.s.s.s nh.nh.nh.nh vrf default.`

This command indicates that packets that are destined to `x.x.x.x/s.s.s.s` are reachable through `nh.nh.nh.nh` in the default VRF table. Meaning, the routes to `x.x.x.x/s.s.s.s` are leaked from the default VRF routing table into the non-default VRF routing table.

The following example illustrates how route leaking between two VRFs can be performed:

```
interface TenGigabitEthernet 0/9
  ip vrf forwarding VRF1
  ip address 120.0.0.1/24
interface TenGigabitEthernet 0/10
  ip vrf forwarding VRF2
  ip address 140.0.0.1/24
ip route vrf VRF1 20.0.0.0/16 140.0.0.2 vrf VRF2
ip route vrf VRF2 40.0.0.0/16 120.0.0.2 vrf VRF1
```

55

Virtual Router Redundancy Protocol (VRRP)

Virtual router redundancy protocol (VRRP) is supported on the S6000 platform.

VRRP Overview

VRRP is designed to eliminate a single point of failure in a statically routed network.

VRRP specifies a MASTER router that owns the next hop IP and MAC address for end stations on a local area network (LAN). The MASTER router is chosen from the virtual routers by an election process and forwards packets sent to the next hop IP address. If the MASTER router fails, VRRP begins the election process to choose a new MASTER router and that new MASTER continues routing traffic.

VRRP uses the virtual router identifier (VRID) to identify each virtual router configured. The IP address of the MASTER router is used as the next hop address for all end stations on the LAN. The other routers the IP addresses represent are BACKUP routers.

VRRP packets are transmitted with the virtual router MAC address as the source MAC address. The MAC address is in the following format: 00-00-5E-00-01-{VRID}. The first three octets are unchangeable. The next two octets (00-01) indicate the address block assigned to the VRRP protocol, and are unchangeable. The final octet changes depending on the VRRP virtual router identifier and allows for up to 255 VRRP routers on a network.

The following example shows a typical network configuration using VRRP. Instead of configuring the hosts on the network 10.10.10.0 with the IP address of either Router A or Router B as their default router; their default router is the IP address configured on the virtual router. When any host on the LAN segment wants to access the Internet, it sends packets to the IP address of the virtual router.

In the following example, Router A is configured as the MASTER router. It is configured with the IP address of the virtual router and sends any packets addressed to the virtual router through interface GigabitEthernet 1/1 to the Internet. As the BACKUP router, Router B is also configured with the IP address of the virtual router. If, for any reason, Router A becomes unavailable, VRRP elects a new MASTER Router. Router B assumes the duties of Router A and becomes the MASTER router. At that time, Router B responds to the packets sent to the virtual IP address.

All workstations continue to use the IP address of the virtual router to address packets destined to the Internet. Router B receives and forwards them on interface GigabitEthernet 10/1. Until Router A resumes operation, VRRP allows Router B to provide uninterrupted service to the users on the LAN segment accessing the Internet.

For more detailed information about VRRP, refer to *RFC 2338, Virtual Router Redundancy Protocol*.

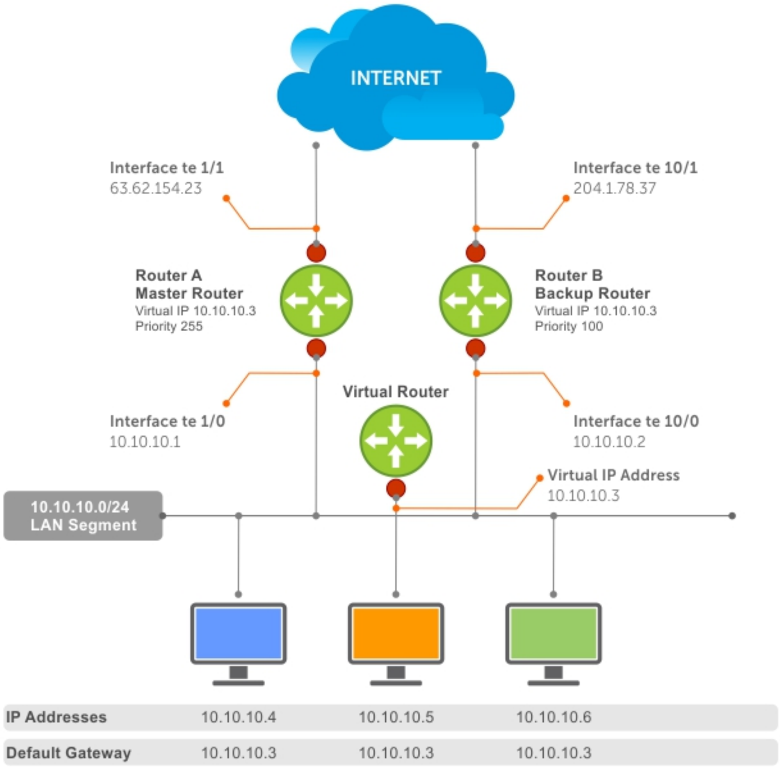

INTERNET
Interface te 1/1
63.62.154.23
Interface te 10/1
204.1.78.37
Router A
Master Router
Virtual IP 10.10.10.3
Priority 255
Router B
Backup Router
Virtual IP 10.10.10.3
Priority 100
Virtual Router
Interface te 1/0
10.10.10.1
Interface te 10/0
10.10.10.2
Virtual IP Address
10.10.10.3
10.10.10.0/24
LAN Segment
IP Addresses
10.10.10.4
10.10.10.5
10.10.10.6
Default Gateway
10.10.10.3
10.10.10.3
10.10.10.3

Figure 125. Basic VRRP Configuration

VRRP Benefits

With VRRP configured on a network, end-station connectivity to the network is not subject to a single point-of-failure. End-station connections to the network are redundant and are not dependent on internal gateway protocol (IGP) protocols to converge or update routing tables.

VRRP Implementation

Within a single VRRP group, up to 12 virtual IP addresses are supported.

Virtual IP addresses can belong to the primary or secondary IP address' subnet configured on the interface. You can ping all the virtual IP addresses configured on the Master VRRP router from anywhere in the local subnet.

The following recommendations shown may vary depending on various factors like address resolution protocol (ARP) broadcasts, IP broadcasts, or spanning tree protocol (STP) before changing the advertisement interval. When the number of packets processed by RP2/CP/FP processor increases or decreases based on the dynamics of the network, the advertisement intervals may increase or decrease accordingly.

CAUTION: Increasing the advertisement interval increases the VRRP Master dead interval, resulting in an increased failover time for Master/Backup election. Take caution when increasing the advertisement interval, as the increased dead interval may cause packets to be dropped during that switch-over time.

VRRP Configuration

By default, VRRP is not configured.

Configuration Task List

The following list specifies the configuration tasks for VRRP.

- Creating a Virtual Router (mandatory)
- Configuring the VRRP Version for an IPv4 Group (optional)
- Assign Virtual IP Addresses mandatory)
- Setting VRRP Group (Virtual Router) Priority (optional)
- Configuring VRRP Authentication (optional)
- Disabling Preempt (optional)
- Changing the Advertisement Interval (optional)
- Track an Interface or Object
- Setting VRRP Initialization Delay

For a complete listing of all commands related to VRRP, refer to *Dell Networking OS Command Line Reference Guide*.

Creating a Virtual Router

To enable VRRP, create a virtual router. In Dell Networking Operating System (OS), the virtual router identifier (VRID) identifies a VRRP group.
To enable or delete a virtual router, use the following commands.

- Create a virtual router for that interface with a VRID.

  INTERFACE mode

  ```
  vrrp-group vrid
  ```

  The VRID range is from 1 to 255.

  NOTE: The interface must already have a primary IP address defined and be enabled, as shown in the second example.
- Delete a VRRP group.

  INTERFACE mode

  ```
  no vrrp-group vrid
  ```

Examples of Configuring and Verifying VRRP

The following examples how to configure VRRP.

```
Dell(conf)#int gi 1/1
Dell(conf-if-gi-1/1)#vrrp-group 111
Dell(conf-if-gi-1/1-vrid-111)#
```

The following examples how to verify the VRRP configuration.

```
Dell(conf-if-gi-1/1)#show conf
!
interface GigabitEthernet 1/1
 ip address 10.10.10.1/24
!
 vrrp-group 111
 no shutdown
Dell(conf-if-gi-1/1)#
```

Configuring the VRRP Version for an IPv4 Group

For IPv4, you can configure a VRRP group to use one of the following VRRP versions:

- VRRPv2 as defined in RFC 3768, *Virtual Router Redundancy Protocol (VRRP)*
- VRRPv3 as defined in RFC 5798, *Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6*

You can also migrate a IPv4 group from VRRPv2 to VRRP3.

To configure the VRRP version for IPv4, use the **version** command in INTERFACE mode.

Example: Configuring VRRP to Use Version 3

The following example configures the IPv4 VRRP 100 group to use VRRP protocol version 3.

```
Dell(conf-if-te-0/0)# vrrp-group 100
Dell (conf-if-te-0/0-vrid-100)#version ?
2      VRRPv2
3      VRRPv3
both   Interoperable, send VRRPv3 receive both
Dell(conf-if-te-0/0-vrid-100)#version 3
```

You can use the **version both** command in INTERFACE mode to migrate from VRRPv2 to VRRPv3. When you set the VRRP version to **both**, the switch sends only VRRPv3 advertisements but can receive VRRPv2 or VRRPv3 packets.

To migrate an IPv4 VRRP group from VRRPv2 to VRRPv3:

1. Set the switches with the lowest priority to “both”.
2. Set the switch with the highest priority to version to 3.
3. Set all the switches from **both** to version 3.

NOTE: Do not run VRRP version 2 and version 3 in the same group for an extended period of time

Example: Migrating an IPv4 VRRP Group from VRRPv2 to VRRPv3

NOTE: Carefully following this procedure, otherwise you might introduce dual master switches issues.

To migrate an IPv4 VRRP Group from VRRPv2 to VRRPv3:

1. Set the backup switches to VRRP version to both.

```
Dell_backup_switch1(conf-if-te-0/1-vrid-100)#version both
Dell_backup_switch2(conf-if-te-0/2-vrid-100)#version both
```

2. Set the master switch to VRRP protocol version 3.

```
Dell_master_switch(conf-if-te-0/1-vrid-100)#version 3
```

3. Set the backup switches to version 3.

```
Dell_backup_switch1(conf-if-te-0/1-vrid-100)#version 3
Dell_backup_switch2(conf-if-te-0/2-vrid-100)#version 3
```

Assign Virtual IP addresses

Virtual routers contain virtual IP addresses configured for that VRRP group (VRID). A VRRP group does not transmit VRRP packets until you assign the Virtual IP address to the VRRP group.

For more information, refer to VRRP Implementation.

To activate a VRRP group on an interface (so that VRRP group starts transmitting VRRP packets), configure at least one virtual IP address in a VRRP group. The virtual IP address is the IP address of the virtual router and does not require the IP address mask.

You can configure up to 12 virtual IP addresses on a single VRRP group (VRID).

The following rules apply to virtual IP addresses:

- The virtual IP addresses must be in the same subnet as the primary or secondary IP addresses configured on the interface. Though a single VRRP group can contain virtual IP addresses belonging to multiple IP subnets configured on the interface, Dell Networking recommends configuring virtual IP addresses belonging to the same IP subnet for any one VRRP group.
  - For example, an interface (on which you enable VRRP) contains a primary IP address of 50.1.1.1/24 and a secondary IP address of 60.1.1.1/24. The VRRP group (VRID 1) must contain virtual addresses belonging to either subnet 50.1.1.0/24 or subnet 60.1.1.0/24, but not from both subnets (though Dell Networking OS allows the same).
- If the virtual IP address and the interface’s primary/secondary IP address are the same, the priority on that VRRP group MUST be set to 255. The interface then becomes the OWNER router of the VRRP group and the interface’s physical MAC address is changed to that of the owner VRRP group’s MAC address.
- If you configure multiple VRRP groups on an interface, only one of the VRRP Groups can contain the interface primary or secondary IP address.

Configuring a Virtual IP Address

To configure a virtual IP address, use the following commands.

1. Configure a VRRP group.

INTERFACE mode

```
vrrp-group vrrp-id
```

The VRID range is from 1 to 255.

2. Configure virtual IP addresses for this VRID.

   INTERFACE -VRID mode

   ```
   virtual-address ip-address1 [...ip-address12]
   ```

   The range is up to 12 addresses.

Examples of the Configuring and Verifying a Virtual IP Address

The following example shows how to configure a virtual IP address.

```
Dell(conf-if-gi-1/1-vrid-111)#virtual-address 10.10.10.1
Dell(conf-if-gi-1/1-vrid-111)#virtual-address 10.10.10.2
Dell(conf-if-gi-1/1-vrid-111)#virtual-address 10.10.10.3
Dell(conf-if-gi-1/1-vrid-111)#
```

The following example shows how to verify a virtual IP address configuration.

NOTE: In the following example, the primary IP address and the virtual IP addresses are on the same subnet.

```
Dell(conf-if-gi-1/1)#show conf
!
interface GigabitEthernet 1/1
  ip address 10.10.10.1/24
!
vrrp-group 111
  priority 255
  virtual-address 10.10.10.1
  virtual-address 10.10.10.2
  virtual-address 10.10.10.3
!
vrrp-group 222
  no shutdown
Dell(conf-if-gi-1/1)#
```

The following example shows the same VRRP group (VRID 111) configured on multiple interfaces on different subnets.

```
Dellshow vrrp
------------------
GigabitEthernet 1/1, VRID: 111, Net: 10.10.10.1
State: Master, Priority: 255, Master: 10.10.10.1 (local)
Hold Down: 0 sec, Preempt: TRUE, AdvInt: 1 sec
Adv rcvd: 0, Bad pkts rcvd: 0, Adv sent: 1768, Gratuitous ARP sent: 5
Virtual MAC address:
  00:00:5e:00:01:6f
Virtual IP address:
  10.10.10.1 10.10.10.2 10.10.10.3 10.10.10.10
Authentication: (none)
------------------
GigabitEthernet 1/2, VRID: 111, Net: 10.10.2.1
State: Master, Priority: 100, Master: 10.10.2.1 (local)
Hold Down: 0 sec, Preempt: TRUE, AdvInt: 1 sec
Adv rcvd: 0, Bad pkts rcvd: 0, Adv sent: 27, Gratuitous ARP sent: 2
Virtual MAC address:
  00:00:5e:00:01:6f
Virtual IP address:
  10.10.2.2 10.10.2.3
Authentication:
```

When the VRRP process completes its initialization, the State field contains either Master or Backup.

Setting VRRP Group (Virtual Router) Priority

Setting a virtual router priority to 255 ensures that router is the "owner" virtual router for the VRRP group. VRRP elects the MASTER router by choosing the router with the highest priority.
The default priority for a virtual router is **100**. The higher the number, the higher the priority. If the MASTER router fails, VRRP begins the election process to choose a new MASTER router based on the next-highest priority.

If two routers in a VRRP group come up at the same time and have the same priority value, the interface's physical IP addresses are used as tie-breakers to decide which is MASTER. The router with the higher IP address becomes MASTER.

To configure the VRRP group's priority, use the following command.

- Configure the priority for the VRRP group.

  INTERFACE -VRID mode

  `priority` *`priority`*

  The range is from 1 to 255.

  The default is **100**.

Examples of the `priority` Command

```
Dell(conf-if-gi-1/2)#vrrp-group 111
Dell(conf-if-gi-1/2-vrid-111)#priority 125
```

To verify the VRRP group priority, use the `show vrrp` command.

```
Dellshow vrrp
------------------
GigabitEthernet 1/1, VRID: 111, Net: 10.10.10.1
State: Master, Priority: 255, Master: 10.10.10.1 (local)
Hold Down: 0 sec, Preempt: TRUE, AdvInt: 1 sec
Adv rcvd: 0, Bad pkts rcvd: 0, Adv sent: 2343, Gratuitous ARP sent: 5
Virtual MAC address:
  00:00:5e:00:01:6f
Virtual IP address:
  10.10.10.1 10.10.10.2 10.10.10.3 10.10.10.10
Authentication: (none)
------------------
GigabitEthernet 1/2, VRID: 111, Net: 10.10.2.1
State: Master, Priority: 125, Master: 10.10.2.1 (local)
Hold Down: 0 sec, Preempt: TRUE, AdvInt: 1 sec
Adv rcvd: 0, Bad pkts rcvd: 0, Adv sent: 601, Gratuitous ARP sent: 2
Virtual MAC address:
  00:00:5e:00:01:6f
Virtual IP address:
  10.10.2.2 10.10.2.3
Authentication: (none)
Dell(conf)#
```

Configuring VRRP Authentication

Simple authentication of VRRP packets ensures that only trusted routers participate in VRRP processes. When you enable authentication, Dell Networking OS includes the password in its VRRP transmission. The receiving router uses that password to verify the transmission.\

NOTE: You must configure all virtual routers in the VRRP group the same: you must enable authentication with the same password or authentication is disabled.

To configure simple authentication, use the following command.

- Configure a simple text password.

  INTERFACE-VRID mode

  ```
  authentication-type simple [encryption-type] password
  ```

  Parameters:

  - *encryption-type*: `0` indicates unencrypted; `7` indicates encrypted.
  - *password*: plain text.

Examples of the `authentication-type` Command

The bold section shows the encryption type (encrypted) and the password.

```
Dell(conf-if-gi-1/1-vrid-111)#authentication-type ?
Dell(conf-if-gi-1/1-vrid-111)#authentication-type simple 7 force10
```

The following example shows verifying the VRRP authentication configuration using the `show conf` command. The bold section shows the encrypted password.

```
Dell(conf-if-gi-1/1-vrid-111)#show conf
!
  vrrp-group 111
    authentication-type simple 7 387a7f2df5969da4
    priority 255
    virtual-address 10.10.10.1
    virtual-address 10.10.10.2
    virtual-address 10.10.10.3
    virtual-address 10.10.10.10
Dell(conf-if-gi-1/1-vrid-111)#
```

Disabling Preempt

The `preempt` command is enabled by default. The command forces the system to change the MASTER router if another router with a higher priority comes online.
Prevent the BACKUP router with the higher priority from becoming the MASTER router by disabling preempt.

NOTE: You must configure all virtual routers in the VRRP group the same: you must configure all with preempt enabled or configure all with preempt disabled.

Because preempt is enabled by default, disable the preempt function with the following command.

- Prevent any BACKUP router with a higher priority from becoming the MASTER router.

  INTERFACE-VRID mode

  ```
  no preempt
  ```

Examples of Disabling Preempt

Re-enable preempt by entering the `preempt` command. When you enable preempt, it does not display in the `show` commands, because it is a default setting.

The following example shows how to disable preempt using the `no preempt` command.

```
Dell(conf-if-gi-1/1)#vrrp-group 111
Dell(conf-if-gi-1/1-vrid-111)#no preempt
Dell(conf-if-gi-1/1-vrid-111)#
```

The following example shows how to verify preempt is disabled using the `show conf` command.

```
Dell(conf-if-gi-1/1-vrid-111)#show conf
!
  vrrp-group 111
    authentication-type simple 7 387a7f2df5969da4
    no preempt
    priority 255
    virtual-address 10.10.10.1
    virtual-address 10.10.10.2
    virtual-address 10.10.10.3
    virtual-address 10.10.10.10
Dell(conf-if-gi-1/1-vrid-111)#
```

Changing the Advertisement Interval

By default, the MASTER router transmits a VRRP advertisement to all members of the VRRP group every one second, indicating it is operational and is the MASTER router.
If the VRRP group misses three consecutive advertisements, the election process begins and the BACKUP virtual router with the highest priority transitions to MASTER.

NOTE: To avoid throttling VRRP advertisement packets, Dell Networking OS recommends increasing the VRRP advertisement interval to a value higher than the default value of one second. If you do change the time interval between VRRP advertisements on one router, change it on all participating routers.

If are using VRRP version 2, you must configure the timer values in multiple of whole seconds. For example a timer value of 3 seconds or 300 centisecs are valid and equivalent. However, a time value of 50 centisecs is invalid because it not a multiple of 1 second. If you are using VRRP version 3, you must configure the timer values in multiples of 25 centisecs.

If you are configured for VRRP version 2, the timer values must be in multiples of whole seconds. For example, timer value of 3 seconds or 300 centisecs are valid and equivalent. However, a timer value of 50 centisecs is invalid because it not is not multiple of 1 second.

If are using VRRP version 3, you must configure the timer values in multiples of 25 centisecs.

To change the advertisement interval in seconds or centisecs, use the following command. A centisecs is 1/100 of a second.

- Change the advertisement interval setting.

  INTERFACE-VRID mode

  `advertise-interval` *`seconds`*

  The range is from 1 to 255 seconds.

  The default is **1 second**.
- For VRRPv3, change the advertisement centisecs interval setting.

  INTERFACE-VRID mode

`advertise-interval centisecs` *`centisecs`*

The range is from 25 to 4075 centisecs in units of 25 centisecs.

The default is 100 centisecs.

Examples of the `advertise-interval` Command

The following example shows how to change the advertise interval using the `advertise-interval` command.

```
Dell(conf-if-gi-1/1)#vrrp-group 111
Dell(conf-if-gi-1/1-vrid-111)#advertise-interval 10
Dell(conf-if-gi-1/1-vrid-111)#
```

The following example shows how to verify the advertise interval change using the `show conf` command.

```
Dell(conf-if-gi-1/1-vrid-111)#show conf
!
  vrrp-group 111
    advertise-interval 10
    authentication-type simple 7 387a7f2df5969da4
    no preempt
    priority 255
    virtual-address 10.10.10.1
    virtual-address 10.10.10.2
    virtual-address 10.10.10.3
    virtual-address 10.10.10.10
Dell(conf-if-gi-1/1-vrid-111)#
```

Track an Interface or Object

You can set Dell Networking OS to monitor the state of any interface according to the virtual group.

Each VRRP group can track up to 12 interfaces and up to 20 additional objects, which may affect the priority of the VRRP group. If the tracked interface goes down, the VRRP group's priority decreases by a default value of **10** (also known as *cost*). If the tracked interface's state goes up, the VRRP group's priority increases by 10.

The lowered priority of the VRRP group may trigger an election. As the Master/Backup VRRP routers are selected based on the VRRP group's priority, tracking features ensure that the best VRRP router is the Master for that group. The sum of all the costs of all the tracked interfaces must be less than the configured priority on the VRRP group. If the VRRP group is configured as Owner router (priority 255), tracking for that group is disabled, irrespective of the state of the tracked interfaces. The priority of the owner group always remains at 255.

For a virtual group, you can track the line-protocol state or the routing status of any of the following interfaces with the `interface` *`interface`* parameter:

- 1 Gigabit Ethernet: enter `gigabitethernet` *`slot/port`* in the track interface command (shown in the following example).
- 10 Gigabit Ethernet: enter `tengigabitethernet` *`slot/port`*.
- Port channel: enter `port-channel` *`number`*.
- SONET: enter `sonet` *`slot/port`*.
- VLAN: enter `vlan` *`vlan-id`*.

  - The valid VLAN IDs are from 1 to 4094.

For a virtual group, you can also track the status of a configured object (the `track` *`object-id`* command) by entering its object number.

NOTE: You can configure a tracked object for a VRRP group (using the `track` *`object-id`* command in INTERFACE-VRID mode) before you actually create the tracked object (using a `track` *`object-id`* command in CONFIGURATION mode). However, no changes in the VRRP group's priority occur until the tracked object is defined and determined to be down.

In addition, if you configure a VRRP group on an interface that belongs to a VRF instance and later configure object tracking on an interface for the VRRP group, the tracked interface must belong to the VRF instance.

Tracking an Interface

To track an interface, use the following commands.

NOTE: The sum of all the costs for all tracked interfaces must be less than the configured priority of the VRRP group.

- Monitor an interface and, optionally, set a value to be subtracted from the interface's VRRP group priority.

  INTERFACE-VRID mode

  ```
  track interface [priority-cost cost]
  ```

  The cost range is from 1 to 254.

  The default is **10**.
- (Optional) Display the configuration and the UP or DOWN state of tracked objects, including the client (VRRP group) that is tracking an object's state.

  EXEC mode or EXEC Privilege mode

  ```
  show track
  ```
- (Optional) Display the configuration and the UP or DOWN state of tracked interfaces and objects in VRRP groups, including the time since the last change in an object's state.

  EXEC mode or EXEC Privilege mode

  ```
  show vrrp
  ```
- (Optional) Display the configuration of tracked objects in VRRP groups on a specified interface.

  EXEC mode or EXEC Privilege mode

  ```
  show running-config interface interface
  ```

Examples of Configuring and Viewing the `track` Command

The following example shows how to configure tracking using the `track` command.

```
Dell(conf-if-gi-1/1)#vrrp-group 111
Dell(conf-if-gi-1/1-vrid-111)#track gigabitethernet 1/2
Dell(conf-if-gi-1/1-vrid-111)#
```

The following example shows how to verify tracking using the `show conf` command.

```
Dell(conf-if-gi-1/1-vrid-111)#show conf
!
```

```
  vrrp-group 111
    advertise-interval 10
    authentication-type simple 7 387a7f2df5969da4
    no preempt
    priority 255
    track GigabitEthernet 1/2
    virtual-address 10.10.10.1
    virtual-address 10.10.10.2
    virtual-address 10.10.10.3
    virtual-address 10.10.10.10
Dell(conf-if-gi-1/1-vrid-111)#
```

The following example shows verifying the tracking status.

```
Dell#show track

Track 2
  IPv6 route 2040::/64 metric threshold
  Metric threshold is Up (STATIC/0/0)
    5 changes, last change 00:02:16
  Metric threshold down 255 up 254
  First-hop interface is GigabitEthernet 13/2
  Tracked by:
    VRRP GigabitEthernet 7/30 IPv6 VRID 1

Track 3
  IPv6 route 2050::/64 reachability
  Reachability is Up (STATIC)
    5 changes, last change 00:02:16
  First-hop interface is GigabitEthernet 13/2
  Tracked by:
    VRRP GigabitEthernet 7/30 IPv6 VRID 1
```

The following example shows verifying the VRRP status.

```
Dell#show vrrp
------------------

GigabitEthernet 7/30, IPv6 VRID: 1, Version: 3, Net: fe80::201:e8ff:fe01:95cc
VRF: 0 default-vrf
State: Master, Priority: 100, Master: fe80::201:e8ff:fe01:95cc (local)
Hold Down: 0 centisec, Preempt: TRUE, AdvInt: 100 centisec
Accept Mode: FALSE, Master AdvInt: 100 centisec
Adv rcvd: 0, Bad pkts rcvd: 0, Adv sent: 310
Virtual MAC address:
  00:00:5e:00:02:01
Virtual IP address:
  2007::1 fe80::1
Tracking states for 2 resource Ids:
 2 - Up IPv6 route, 2040::/64, priority-cost 20, 00:02:11
 3 - Up IPv6 route, 2050::/64, priority-cost 30, 00:02:11
```

The following example shows verifying the VRRP configuration on an interface.

```
Dell#show running-config interface gigabitethernet 7/30

interface GigabitEthernet 7/30
  no ip address
  ipv6 address 2007::30/64

vrrp-ipv6-group 1
 track 2 priority-cost 20
 track 3 priority-cost 30
```

```
  virtual-address 2007::1
  virtual-address fe80::1
no shutdown
```

Setting VRRP Initialization Delay

VRRP initialization delay is supported on the S6000 platform.
When configured, VRRP is enabled immediately upon system reload or boot. You can delay VRRP initialization to allow the IGP and EGP protocols to be enabled prior to selecting the VRRP Master. This delay ensures that VRRP initializes with no errors or conflicts. You can configure the delay for up to 15 minutes, after which VRRP enables normally.

NOTE: When you reload a node that contains VRRP configuration and is enabled for VLT, Dell Networking recommends that you configure the reload timer by using the `vrrp delay reload` command to ensure that VRRP is functional. Otherwise, when you reload a VLT node configured for VRRP, the local destination address is not seen on the reloaded node causing suboptimal routing.

Set the delay timer on individual interfaces. The delay timer is supported on all physical interfaces, VLANs, and LAGs.

When you configure both CLIs, the later timer rules VRRP enabling. For example, if you set `vrrp delay reload 600` and `vrrp delay minimum 300`, the following behavior occurs:

- When the system reloads, VRRP waits 600 seconds (10 minutes) to bring up VRRP on all interfaces that are up and configured for VRRP.
- When an interface comes up and becomes operational, the system waits 300 seconds (5 minutes) to bring up VRRP on that interface.

To set the delay time for VRRP initialization, use the following commands.

- Set the delay time for VRRP initialization on an individual interface.

  INTERFACE mode

  `vrrp delay minimum` *`seconds`*

  This time is the gap between an interface coming up and being operational, and VRRP enabling.

  The seconds range is from 0 to 900.

  The default is **0**.
- Set the delay time for VRRP initialization on all the interfaces in the system configured for VRRP.

  INTERFACE mode

  `vrrp delay reload` *`seconds`*

  This time is the gap between system boot up completion and VRRP enabling.

  The seconds range is from 0 to 900.

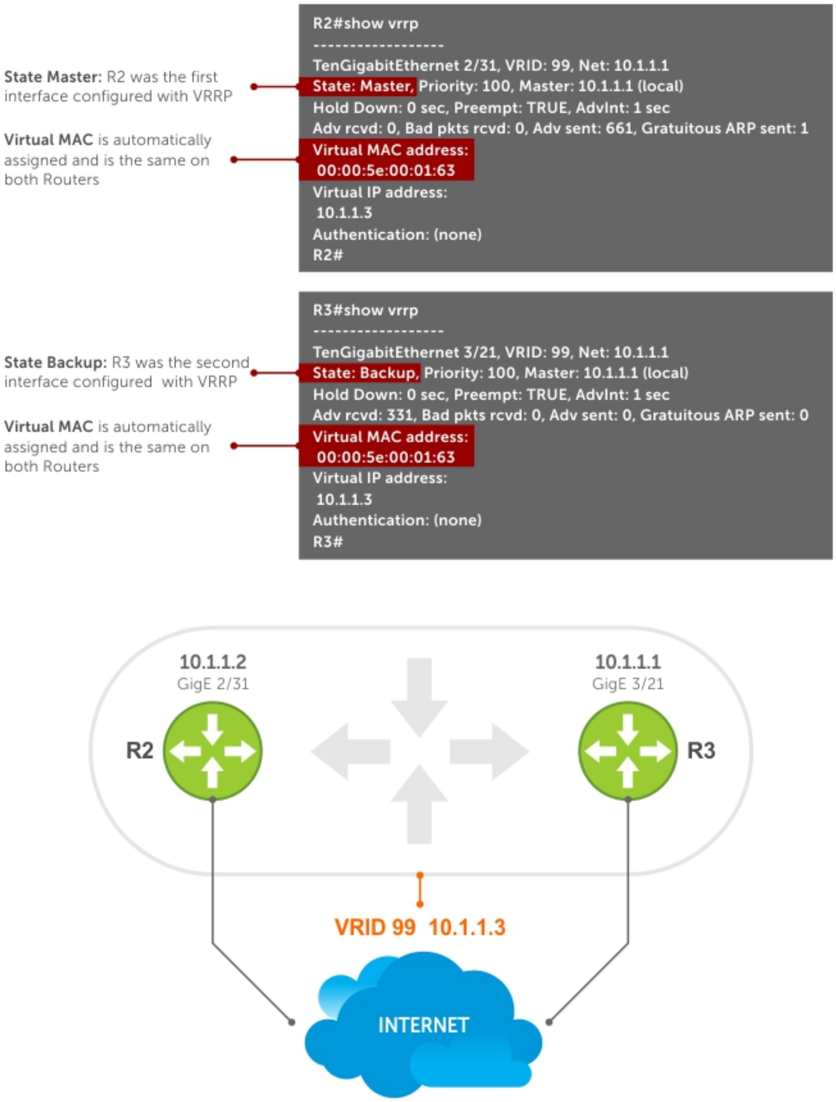
  The default is **0**.

Sample Configurations

Before you set up VRRP, review the following sample configurations.

VRRP for an IPv4 Configuration

The following configuration shows how to enable IPv4 VRRP. This example does not contain comprehensive directions and is intended to provide guidance for only a typical VRRP configuration. You can copy and paste from the example to your CLI. To support your own IP addresses, interfaces, names, and so on, be sure that you make the necessary changes. The VRRP topology was created using the CLI configuration shown in the following example.



State Master: R2 was the first interface configured with VRRP

Virtual MAC is automatically assigned and is the same on both Routers

```
R2#show vrrp
------------------
TenGigabitEthernet 2/31, VRID: 99, Net: 10.1.1.1
State: Master, Priority: 100, Master: 10.1.1.1 (local)
Hold Down: 0 sec, Preempt: TRUE, AdvInt: 1 sec
Adv rcvd: 0, Bad pkts rcvd: 0, Adv sent: 661, Gratuitous ARP sent: 1
Virtual MAC address:
 00:00:5e:00:01:63
Virtual IP address:
 10.1.1.3
Authentication: (none)
R2#
```

State Backup: R3 was the second interface configured with VRRP

Virtual MAC is automatically assigned and is the same on both Routers

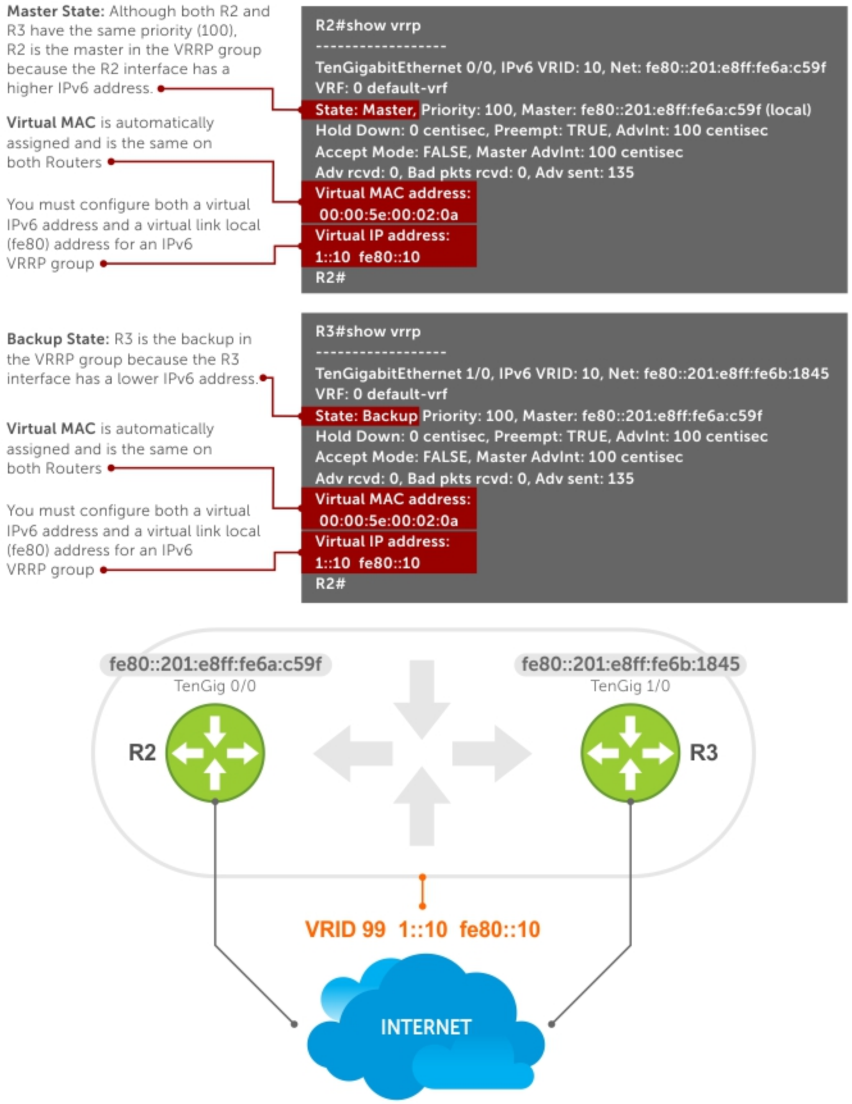
```
R3#show vrrp
------------------
TenGigabitEthernet 3/21, VRID: 99, Net: 10.1.1.1
State: Backup, Priority: 100, Master: 10.1.1.1 (local)
Hold Down: 0 sec, Preempt: TRUE, AdvInt: 1 sec
Adv rcvd: 331, Bad pkts rcvd: 0, Adv sent: 0, Gratuitous ARP sent: 0
Virtual MAC address:
 00:00:5e:00:01:63
Virtual IP address:
 10.1.1.3
Authentication: (none)
R3#
```

Figure 126. VRRP for IPv4 Topology

Examples of Configuring VRRP for IPv4 and IPv6

The following example shows configuring VRRP for IPv4 Router 2.

```
R2(conf)#int gi 2/31
R2(conf-if-gi-2/31)#ip address 10.1.1.1/24
R2(conf-if-gi-2/31)#vrrp-group 99
R2(conf-if-gi-2/31-vrid-99)#priority 200
R2(conf-if-gi-2/31-vrid-99)#virtual 10.1.1.3
R2(conf-if-gi-2/31-vrid-99)#no shut
R2(conf-if-gi-2/31)#show conf
!
interface GigabitEthernet 2/31
  ip address 10.1.1.1/24
!
  vrrp-group 99
    priority 200
    virtual-address 10.1.1.3
  no shutdown
R2(conf-if-gi-2/31)#end

R2#show vrrp
------------------
GigabitEthernet 2/31, VRID: 99, Net: 10.1.1.1
State: Master, Priority: 200, Master: 10.1.1.1 (local)
Hold Down: 0 sec, Preempt: TRUE, AdvInt: 1 sec
Adv rcvd: 0, Bad pkts rcvd: 0, Adv sent: 817, Gratuitous ARP sent: 1
Virtual MAC address:
  00:00:5e:00:01:63
Virtual IP address:
  10.1.1.3
Authentication: (none)
R2#
Router 3
R3(conf)#int gi 3/21
R3(conf-if-gi-3/21)#ip address 10.1.1.2/24
R3(conf-if-gi-3/21)#vrrp-group 99
R3(conf-if-gi-3/21-vrid-99)#virtual 10.1.1.3
R3(conf-if-gi-3/21-vrid-99)#no shut
R3(conf-if-gi-3/21)#show conf
!
interface GigabitEthernet 3/21
  ip address 10.1.1.1/24
!
  vrrp-group 99
   virtual-address 10.1.1.3
  no shutdown
R3(conf-if-gi-3/21)#end
R3#show vrrp
------------------
GigabitEthernet 3/21, VRID: 99, Net: 10.1.1.2
State: Backup, Priority: 100, Master: 10.1.1.1
Hold Down: 0 sec, Preempt: TRUE, AdvInt: 1 sec
Adv rcvd: 698, Bad pkts rcvd: 0, Adv sent: 0, Gratuitous ARP sent: 0
Virtual MAC address:
  00:00:5e:00:01:63
Virtual IP address:
  10.1.1.3
Authentication: (none)
```



Master State: Although both R2 and R3 have the same priority (100), R2 is the master in the VRRP group because the R2 interface has a higher IPv6 address.

Virtual MAC is automatically assigned and is the same on both Routers

You must configure both a virtual IPv6 address and a virtual link local (fe80) address for an IPv6 VRRP group

```
R2#show vrrp
------------------
TenGigabitEthernet 0/0, IPv6 VRID: 10, Net: fe80::201:e8ff:fe6a:c59f
VRF: 0 default-vrf
State: Master, Priority: 100, Master: fe80::201:e8ff:fe6a:c59f (local)
Hold Down: 0 centisec, Preempt: TRUE, AdvInt: 100 centisec
Accept Mode: FALSE, Master AdvInt: 100 centisec
Adv rcvd: 0, Bad pkts rcvd: 0, Adv sent: 135
Virtual MAC address:
 00:00:5e:00:02:0a
Virtual IP address:
1::10 fe80::10
R2#
```

Backup State: R3 is the backup in the VRRP group because the R3 interface has a lower IPv6 address.

Virtual MAC is automatically assigned and is the same on both Routers

You must configure both a virtual IPv6 address and a virtual link local (fe80) address for an IPv6 VRRP group

```
R3#show vrrp
------------------
TenGigabitEthernet 1/0, IPv6 VRID: 10, Net: fe80::201:e8ff:fe6b:1845
VRF: 0 default-vrf
State: Backup Priority: 100, Master: fe80::201:e8ff:fe6a:c59f
Hold Down: 0 centisec, Preempt: TRUE, AdvInt: 100 centisec
Accept Mode: FALSE, Master AdvInt: 100 centisec
Adv rcvd: 0, Bad pkts rcvd: 0, Adv sent: 135
Virtual MAC address:
 00:00:5e:00:02:0a
Virtual IP address:
1::10 fe80::10
R2#
```

Figure 127. VRRP for an IPv6 Configuration

NOTE: In a VRRP or VRRPv3 group, if two routers come up with the same priority and another router already has MASTER status, the router with master status continues to be MASTER even if one of two routers has a higher IP or IPv6 address.

The following example shows configuring VRRP for IPv6 Router 2 and Router 3.

Configure a virtual link local (fe80) address for each VRRPv3 group created for an interface. The VRRPv3 group becomes active as soon as you configure the link local address. Afterward, you can configure the group's virtual IPv6 address.

The virtual IPv6 address you configure must be the same as the IPv6 subnet to which the interface belongs.

Although R2 and R3 have the same default, priority (100), R2 is elected master in the VRRPv3 group because the GigE 0/0 interface has a higher IPv6 address than the GigE 1/0 interface on R3.

Router 2

```
R2(conf)#interface gigabitethernet 0/0
R2(conf-if-gi-0/0)#no ip address
R2(conf-if-gi-0/0)#ipv6 address 1::1/64
R2(conf-if-gi-0/0)#vrrp-group 10
R2(conf-if-gi-0/0-vrid-10)#virtual-address fe80::10
R2(conf-if-gi-0/0-vrid-10)#virtual-address 1::10
R2(conf-if-gi-0/0-vrid-10)#no shutdown
R2(conf-if-gi-0/0)#show config
interface GigabitEthernet 0/0
  ipv6 address 1::1/64
  vrrp-group 10
    priority 100
    virtual-address fe80::10
    virtual-address 1::10
  no shutdown
R2(conf-if-gi-0/0)#end

R2#show vrrp
------------------
GigabitEthernet 0/0, IPv6 VRID: 10, Version: 3, Net:fe80::201:e8ff:fe6a:c59f
VRF: 0 default-vrf
State: Master, Priority: 100, Master: fe80::201:e8ff:fe6a:c59f (local)
Hold Down: 0 centisec, Preempt: TRUE, AdvInt: 100 centisec
Accept Mode: FALSE, Master AdvInt: 100 centisec
Adv rcvd: 0, Bad pkts rcvd: 0, Adv sent: 135
Virtual MAC address:
00:00:5e:00:02:0a
Virtual IP address:
1::10 fe80::10
```

Router 3

```
R3(conf)#interface gigabitethernet 1/0
R3(conf-if-gi-1/0)#no ipv6 address
R3(conf-if-gi-1/0)#ipv6 address 1::2/64
R3(conf-if-gi-1/0)#vrrp-group 10
R2(conf-if-gi-1/0-vrid-10)#virtual-address fe80::10
R2(conf-if-gi-1/0-vrid-10)#virtual-address 1::10
R3(conf-if-gi-1/0-vrid-10)#no shutdown
R3(conf-if-gi-1/0)#show config
interface GigabitEthernet 1/0
  ipv6 address 1::2/64
  vrrp-group 10
    priority 100
    virtual-address fe80::10
    virtual-address 1::10
  no shutdown
R3(conf-if-gi-1/0)#end

R3#show vrrp
------------------
GigabitEthernet 1/0, IPv6 VRID: 10, Version: 3, Net:
fe80::201:e8ff:fe6b:1845
VRF: 0 default-vrf
State: Backup, Priority: 100, Master: fe80::201:e8ff:fe6a:c59f
Hold Down: 0 centisec, Preempt: TRUE, AdvInt: 100 centisec
Accept Mode: FALSE, Master AdvInt: 100 centisec
Adv rcvd: 11, Bad pkts rcvd: 0, Adv sent: 0
Virtual MAC address:
00:00:5e:00:02:0a
```

VRRP in a VRF Configuration

The following example shows how to enable VRRP operation in a VRF virtualized network for the following scenarios.

- Multiple VRFs on physical interfaces running VRRP.
- Multiple VRFs on VLAN interfaces running VRRP.

To view a VRRP in a VRF configuration, use the `show` commands.

VRRP in a VRF: Non-VLAN Scenario

The following example shows how to enable VRRP in a non-VLAN.

The following example shows a typical use case in which you create three virtualized overlay networks by configuring three VRFs in two E-Series switches. The default gateway to reach the internet in each VRF is a static route with the next hop being the virtual IP address configured in VRRP. In this scenario, a single VLAN is associated with each VRF.

Both Switch-1 and Switch-2 have three VRF instances defined: VRF-1, VRF-2, and VRF-3. Each VRF has a separate physical interface to a LAN switch and an upstream VPN interface to connect to the Internet. Both Switch-1 and Switch-2 use VRRP groups on each VRF instance in order that there is one MASTER and one backup router for each VRF. In VRF-1 and VRF-2, Switch-2 serves as owner-master of the VRRP group and Switch-1 serves as the backup. On VRF-3, Switch-1 is the owner-master and Switch-2 is the backup.

In VRF-1 and VRF-2 on Switch-2, the virtual IP and node IP address, subnet, and VRRP group are the same. On Switch-1, the virtual IP address, subnet, and VRRP group are the same in VRF-1 and VRF-2, but the IP address of the node interface is unique. There is no requirement for the virtual IP and node IP addresses to be the same in VRF-1 and VRF-2; similarly, there is no requirement for the IP addresses to be different. In VRF-3, the node IP addresses and subnet are unique.

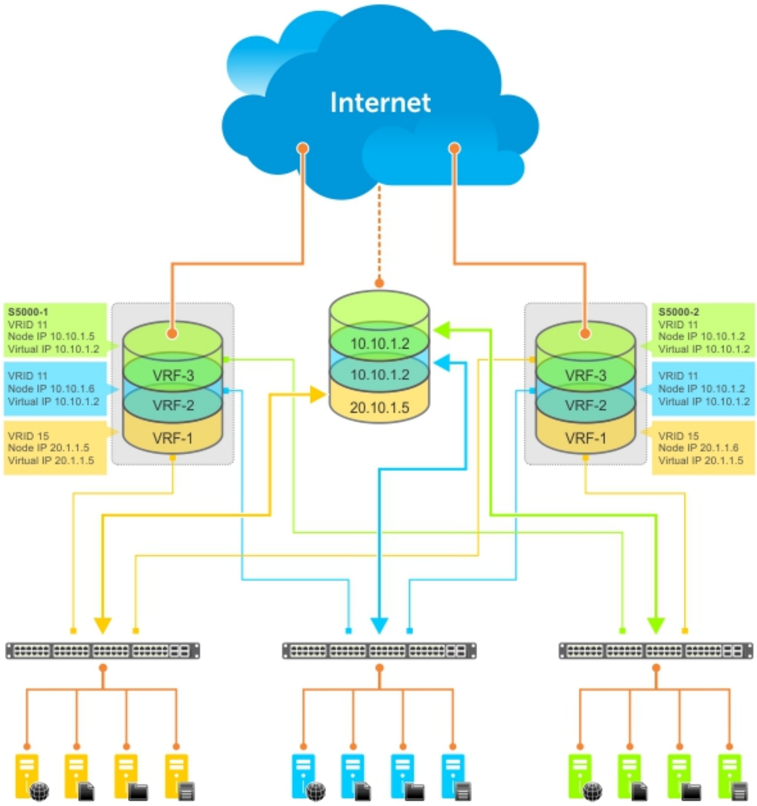

Internet
S5000-1
VRID 11
Node IP 10.10.1.5
Virtual IP 10.10.1.2
VRID 11
Node IP 10.10.1.6
Virtual IP 10.10.1.2
VRID 15
Node IP 20.1.1.5
Virtual IP 20.1.1.5
VRF-3
VRF-2
VRF-1
10.10.1.2
10.10.1.2
20.10.1.5
S5000-2
VRID 11
Node IP 10.10.1.2
Virtual IP 10.10.1.2
VRID 11
Node IP 10.10.1.2
Virtual IP 10.10.1.2
VRID 15
Node IP 20.1.1.6
Virtual IP 20.1.1.5

Figure 128. VRRP in a VRF: Non-VLAN Example

Example of Configuring VRRP in a VRF on Switch-1 (Non-VLAN)

Switch-1

```
S1(conf)#ip vrf default-vrf 0
!
S1(conf)#ip vrf VRF-1 1
!
S1(conf)#ip vrf VRF-2 2
!
S1(conf)#ip vrf VRF-3 3
!
S1(conf)#interface GigabitEthernet 12/1
S1(conf-if-gi-12/1)#ip vrf forwarding VRF-1
S1(conf-if-gi-12/1)#ip address 10.10.1.5/24
S1(conf-if-gi-12/1)#vrrp-group 11
% Info: The VRID used by the VRRP group 11 in VRF 1 will be 177.
S1(conf-if-gi-12/1-vrid-101)#priority 100
S1(conf-if-gi-12/1-vrid-101)#virtual-address 10.10.1.2
S1(conf-if-gi-12/1)#no shutdown
!
S1(conf)#interface GigabitEthernet 12/2
S1(conf-if-gi-12/2)#ip vrf forwarding VRF-2
S1(conf-if-gi-12/2)#ip address 10.10.1.6/24
S1(conf-if-gi-12/2)#vrrp-group 11
% Info: The VRID used by the VRRP group 11 in VRF 2 will be 178.
S1(conf-if-gi-12/2-vrid-101)#priority 100
S1(conf-if-gi-12/2-vrid-101)#virtual-address 10.10.1.2
S1(conf-if-gi-12/2)#no shutdown
```

```
!
S1(conf)#interface GigabitEthernet 12/3
S1(conf-if-gi-12/3)#ip vrf forwarding VRF-3
S1(conf-if-gi-12/3)#ip address 20.1.1.5/24
S1(conf-if-gi-12/3)#vrrp-group 15
% Info: The VRID used by the VRRP group 15 in VRF 3 will be 243.
S1(conf-if-gi-12/3-vrid-105)#priority 255
S1(conf-if-gi-12/3-vrid-105)#virtual-address 20.1.1.5
S1(conf-if-gi-12/3)#no shutdown
```

Example of Configuring VRRP in a VRF on Switch-2 (Non-VLAN Configuration)

Switch-2

```
S2(conf)#ip vrf default-vrf 0
!
S2(conf)#ip vrf VRF-1 1
!
S2(conf)#ip vrf VRF-2 2
!
S2(conf)#ip vrf VRF-3 3
!
S2(conf)#interface GigabitEthernet 12/1
S2(conf-if-gi-12/1)#ip vrf forwarding VRF-1
S2(conf-if-gi-12/1)#ip address 10.10.1.2/24
S2(conf-if-gi-12/1)#vrrp-group 11
% Info: The VRID used by the VRRP group 11 in VRF 1 will be 177.
S2(conf-if-gi-12/1-vrid-101)#priority 255
S2(conf-if-gi-12/1-vrid-101)#virtual-address 10.10.1.2
S2(conf-if-gi-12/1)#no shutdown
!
S2(conf)#interface GigabitEthernet 12/2
S2(conf-if-gi-12/2)#ip vrf forwarding VRF-2
S2(conf-if-gi-12/2)#ip address 10.10.1.2/24
S2(conf-if-gi-12/2)#vrrp-group 11
% Info: The VRID used by the VRRP group 11 in VRF 2 will be 178.
S2(conf-if-gi-12/2-vrid-101)#priority 255
S2(conf-if-gi-12/2-vrid-101)#virtual-address 10.10.1.2
S2(conf-if-gi-12/2)#no shutdown
!
S2(conf)#interface GigabitEthernet 12/3
S2(conf-if-gi-12/3)#ip vrf forwarding VRF-3
S2(conf-if-gi-12/3)#ip address 20.1.1.6/24
S2(conf-if-gi-12/3)#vrrp-group 15
% Info: The VRID used by the VRRP group 15 in VRF 3 will be 243.
S2(conf-if-gi-12/3-vrid-105)#priority 100
S2(conf-if-gi-12/3-vrid-105)#virtual-address 20.1.1.5
S2(conf-if-gi-12/3)#no shutdown
```

VLAN Scenario

In another scenario, to connect to the LAN, VRF-1, VRF-2, and VRF-3 use a single physical interface with multiple tagged VLANs (instead of separate physical interfaces).

In this case, you configure three VLANs: VLAN-100, VLAN-200, and VLAN-300. Each VLAN is a member of one VRF. A physical interface (gigabitethernet 0/1) attaches to the LAN and is configured as a tagged interface in VLAN-100, VLAN-200, and VLAN-300. The rest of this example is similar to the non-VLAN scenario.

This VLAN scenario often occurs in a service-provider network in which you configure VLAN tags for traffic from multiple customers on customer-premises equipment (CPE), and separate VRF instances associated with each VLAN are configured on the provider edge (PE) router in the point-of-presence (POP).

VRRP in VRF: Switch-1 VLAN Configuration

VRRP in VRF: Switch-2 VLAN Configuration

Switch-1

```
S1(conf)#ip vrf VRF-1 1
!
S1(conf)#ip vrf VRF-2 2
!
S1(conf)#ip vrf VRF-3 3
!
S1(conf)#interface GigabitEthernet 12/4
S1(conf-if-gi-12/4)#no ip address
S1(conf-if-gi-12/4)#switchport
S1(conf-if-gi-12/4)#no shutdown
!
S1(conf-if-gi-12/4)#interface vlan 100
S1(conf-if-vl-100)#ip vrf forwarding VRF-1
S1(conf-if-vl-100)#ip address 10.10.1.5/24
S1(conf-if-vl-100)#tagged gigabitethernet 12/4
S1(conf-if-vl-100)#vrrp-group 11
% Info: The VRID used by the VRRP group 11 in VRF 1 will be 177.
S1(conf-if-vl-100-vrid-101)#priority 100
S1(conf-if-vl-100-vrid-101)#virtual-address 10.10.1.2
S1(conf-if-vl-100)#no shutdown
!
S1(conf-if-gi-12/4)#interface vlan 200
S1(conf-if-vl-200)#ip vrf forwarding VRF-2
S1(conf-if-vl-200)#ip address 10.10.1.6/24
S1(conf-if-vl-200)#tagged gigabitethernet 12/4
S1(conf-if-vl-200)#vrrp-group 11
% Info: The VRID used by the VRRP group 11 in VRF 2 will be 178.
S1(conf-if-vl-200-vrid-101)#priority 100
S1(conf-if-vl-200-vrid-101)#virtual-address 10.10.1.2
S1(conf-if-vl-200)#no shutdown
!
S1(conf-if-gi-12/4)#interface vlan 300
S1(conf-if-vl-300)#ip vrf forwarding VRF-3
S1(conf-if-vl-300)#ip address 20.1.1.5/24
S1(conf-if-vl-300)#tagged gigabitethernet 12/4
S1(conf-if-vl-300)#vrrp-group 15
% Info: The VRID used by the VRRP group 15 in VRF 3 will be 243.
S1(conf-if-vl-300-vrid-101)#priority 255
S1(conf-if-vl-300-vrid-101)#virtual-address 20.1.1.5
S1(conf-if-vl-300)#no shutdown
```

Switch-2

```
S2(conf)#ip vrf VRF-1 1
!
S2(conf)#ip vrf VRF-2 2
!
S2(conf)#ip vrf VRF-3 3
!
S2(conf)#interface GigabitEthernet 12/4
S2(conf-if-gi-12/4)#no ip address
S2(conf-if-gi-12/4)#switchport
S2(conf-if-gi-12/4)#no shutdown
!
S2(conf-if-gi-12/4)#interface vlan 100
S2(conf-if-vl-100)#ip vrf forwarding VRF-1
S2(conf-if-vl-100)#ip address 10.10.1.2/24
S2(conf-if-vl-100)#tagged gigabitethernet 12/4
S2(conf-if-vl-100)#vrrp-group 11
% Info: The VRID used by the VRRP group 11 in VRF 1 will be 177.
```

```
S2(conf-if-vl-100-vrid-101)#priority 255
S2(conf-if-vl-100-vrid-101)#virtual-address 10.10.1.2
S2(conf-if-vl-100)#no shutdown
!
S2(conf-if-gi-12/4)#interface vlan 200
S2(conf-if-vl-200)#ip vrf forwarding VRF-2
S2(conf-if-vl-200)#ip address 10.10.1.2/24
S2(conf-if-vl-200)#tagged gigabitethernet 12/4
S2(conf-if-vl-200)#vrrp-group 11
% Info: The VRID used by the VRRP group 11 in VRF 2 will be 178.
S2(conf-if-vl-200-vrid-101)#priority 255
S2(conf-if-vl-200-vrid-101)#virtual-address 10.10.1.2
S2(conf-if-vl-200)#no shutdown
!
S2(conf-if-gi-12/4)#interface vlan 300
S2(conf-if-vl-300)#ip vrf forwarding VRF-3
S2(conf-if-vl-300)#ip address 20.1.1.6/24
S2(conf-if-vl-300)#tagged gigabitethernet 12/4
S2(conf-if-vl-300)#vrrp-group 15
% Info: The VRID used by the VRRP group 15 in VRF 3 will be 243.
S2(conf-if-vl-300-vrid-101)#priority 100
S2(conf-if-vl-300-vrid-101)#virtual-address 20.1.1.5
S2(conf-if-vl-300)#no shutdown
```

56

S-Series Debugging and Diagnostics

This chapter describes debugging and diagnostics for the S6000 platform.

Offline Diagnostics

The offline diagnostics test suite is useful for isolating faults and debugging hardware.

The diagnostics tests are grouped into three levels:

- **Level 0** — Level 0 diagnostics check for the presence of various components and perform essential path verifications. In addition, Level 0 diagnostics verify the identification registers of the components on the board.
- **Level 1** — A smaller set of diagnostic tests. Level 1 diagnostics perform status/self-test for all the components on the board and test their registers for appropriate values. In addition, Level 1 diagnostics perform extensive tests on memory devices (for example, SDRAM, flash, NVRAM, EEPROM) wherever possible.
- **Level 2** — The full set of diagnostic tests. Level 2 diagnostics are used primarily for on-board loopback tests and more extensive component diagnostics. Various components on the board are put into Loopback mode and test packets are transmitted through those components. Level 2 diagnostics also perform snake tests using virtual local area network (VLAN) configurations.

Important Points to Remember

- You can only perform offline diagnostics on an offline standalone unit or offline member unit of a stack of three or more. You cannot perform diagnostics on the management or standby unit in a stack of two or more; if you do, a message similar to this displays: `Running Diagnostics on master/standby unit is not allowed on stack`.
- Perform offline diagnostics on one stack member at a time.
- Diagnostics only test connectivity, not the entire data path.
- Diagnostic results are stored on the flash of the unit on which you performed the diagnostics.
- When offline diagnostics are complete, the unit or stack member reboots automatically.

Running Offline Diagnostics

To run offline diagnostics, use the following commands.
For more information, refer to the examples following the steps.

1. Place the unit in the offline state.
   EXEC Privilege mode

   ```
   offline stack-unit id
   ```

   You cannot enter this command on a MASTER or Standby stack unit.

NOTE: The system reboots when the offline diagnostics complete. This is an automatic process. The following warning message appears when you implement the `offline stack-unit` command: `Warning - Diagnostic execution will cause stack-unit to reboot after completion of diags. Proceed with Offline-Diags [confirm yes/no]:y`

After the system goes offline, you must reload or execute the **online stack-unit <id>** command for the normal operation.

2. Confirm the offline status.
   EXEC Privilege mode

   ```
   show system brief
   ```
3. Start diagnostics on the unit.

   ```
   diag stack-unit id
   ```

   When the tests are complete, the system displays the following message and automatically reboots the unit.

   ```
   Dell#00:09:42 : Diagnostic test results are stored on file: flash:/
   TestReport-SU-0.txt
   Diags completed... Rebooting the system now!!!
   Mar 12 10:40:35: %S6000:0 %DIAGAGT-6-DA_DIAG_DONE: Diags finished on stack
   unit 0
   ```

   Diagnostic results are printed to a file in the flash using the filename format *TestReport-SU-<stack-unit-id>.txt*.

   Log messages differ somewhat when diagnostics are done on a standalone unit and on a stack member.
4. View the results of the diagnostic tests.
   EXEC Privilege mode

   ```
   show file flash://TestReport-SU-stack-unit-id.txt
   ```

Examples of Running Offline Diagnostics

The following example shows the `offline stack-unit id` command.

```
Dell#offline stack-unit 0
Warning - offline of unit will bring down all the protocols and
the unit will be operationally down, except for running Diagnostics.
Please make sure that stacking is not configured for Diagnostics execution.
Also reboot/online command is necessary for normal operation after the offline
command is issued.
Proceed with Offline [confirm yes/no]:yes
Dell#Mar 12 10:32:18: %STKUNIT0-M:CP %CHMGR-2-STACKUNIT_DOWN: Stack unit 0 down
- stack unit offline
Mar 12 10:32:18: %STKUNIT0-M:CP %IFMGR-1-DEL_PORT: Removed port: Te
0/0-3,16-27,32-55,64-67,80-83,88-99,104-107,112-123, Fo
0/4,8,12,28,56,60,68,72,76,84,100,108,124,
```

The following example shows the `show system brief` command.

```
Dell#show system brief

Stack MAC : 10:11:12:13:14:15
Reload-Type                 :   normal-reload [Next boot : normal-reload]
```

```
--  Stack Info  --
Unit  UnitType     Status         ReqTyp          CurTyp          Version
Ports
--------------------------------------------------------------------------------
----
  0   Management   offline         S6000           S6000           9.4(0.0)  128
  1   Member       not present
  2   Member       not present
  3   Member       not present
  4   Member       not present
  5   Member       not present

--  Power Supplies  --
Unit   Bay   Status       Type    FanStatus   FanSpeed(rpm)
---------------------------------------------------------------------------
  0     0    up           AC      up          18048
  0     1    absent               absent      0

--  Fan  Status  --
Unit Bay   TrayStatus  Fan0    Speed   Fan1    Speed
--------------------------------------------------------------------------------
----
 0    0     up          up      19275   up      19275
 0    1     up          up      19275   up      19275
 0    2     absent

Speed in RPM
```

The following example shows the `diag` command (standalone unit).

```
Dell#diag stack-unit 0 level0
Warning - diagnostic execution will cause multiple link flaps on the peer side
- advisable to shut directly connected ports
Proceed with Diags [confirm yes/no]: yes

Dell#Dec 15 04:14:07: %S4820:0 %DIAGAGT-6-DA_DIAG_STARTED: Starting diags on
stack unit 0
00:12:10 : System may take additional time for Driver Init.
00:12:10 : Approximate time to complete the Diags ... 6 Mins
```

The following example shows the `diag` command (stack member).

```
[output from master unit]

Dell#diag stack-unit 2
Warning - the stack unit will be pulled out of the stack for diagnostic
execution
Proceed with Diags [confirm yes/no]: yes
Warning - diagnostic execution will cause multiple link flaps on the peer side
- advisable
to shut directly connected ports
Proceed with Diags [confirm yes/no]: yes
Dell#00:03:13: %S25P:2 %DIAGAGT-6-DA_DIAG_STARTED: Starting diags on stack unit
2
00:03:13 : Approximate time to complete these Diags ... 6 Min
00:03:13 : Diagnostic test results will be stored on stack unit 2 file: flash:/
TestReport-SU-2.txt

Dell#00:03:35: %STKUNIT1-M:CP %CHMGR-2-STACKUNIT_DOWN: Stack unit 2 down - card
removed
00:08:50: %STKUNIT1-M:CP %CHMGR-5-STACKUNITDETECTED: Stack unit 2 present
00:09:00: %STKUNIT1-M:CP %CHMGR-5-CHECKIN: Checkin from Stack unit 2 (type
S25P, 28 ports)
00:09:00: %S25P:2 %CHMGR-0-PS_UP: Power supply 0 in unit 2 is up
00:09:00: %STKUNIT1-M:CP %CHMGR-5-STACKUNITUP: Stack unit 2 is up
[output from the console of the unit in which diagnostics are performed]

Dell(stack-member-2)#
Diagnostic test results are stored on file: flash:/TestReport-SU-2.txt
Diags completed... Rebooting the system now!!!
```

The following example shows the `show file flash:\\` command (standalone member).

```
****************************************************************
Dell# show file flash://TestReport-SU-0.txt
diagS6000CpldCacheDisable[163]: ERROR: platform cpld cache disabled ioctl
failed, rv: 9
*******************************S6000 DIAGNOSTICS*******************************

           Board                  : S6000 Dell Inc.
           CPU Version            : Intel Centerton Processor
           Stack Unit Board Temp  : 32 Degree C
           Stack Unit Number      : 0
           Board Service Tag      : STG1234
           System Cpld Rev        : 0xa
           Master Cpld Rev        : 0xc
           Slave Cpld Rev         : 0xa
           Image Build Version    : 9-4(0-91)

**************************** S6000 LEVEL 0 DIAGNOSTICS*************************

Test 1 - Board Revision Test ........................................ PASS
Test 2 - Cpu Type Detect Test ....................................... PASS
Test 3.000 - Psu0 Presence Test ..................................... PASS
diagS6000PsuPresenceTest[1022]: ERROR: Psu:1 is not present...
Test 3.001 - Psu1 Presence Test ..................................... NOT
PRESENT
Test 3 - Psu Presence Test .......................................... NOT
PRESENT
Test 4.000 - Psu0 Source Type Test .................................. PASS
diagS6000IsPsuGood[954]: ERROR: Psu:1, Power supply is not present.
Test 4.001 - Psu1 Source Type Test .................................. NOT
PRESENT
Test 4 - Psu Source Type Test ....................................... NOT
PRESENT
Test 5.000 - Psu0 Status Monitor Test ............................... PASS
diagS6000PsuStatusMonitorTest[1099]: ERROR: Psu:1, It is not present...
Test 5.001 - Psu1 Status Monitor Test ............................... NOT
PRESENT
Test 5 - Psu Status Monitor Test .................................... NOT
PRESENT
Test 6.000 - Psu0 Fan Speed Monitor Test ............................ PASS
diagS6000IsPsuGood[954]: ERROR: Psu:1, Power supply is not present.
Test 6.001 - Psu1 Fan Speed Monitor Test ............................ NOT
PRESENT
Test 6 - Psu Fan Speed Monitor Test ................................. NOT
PRESENT
Test 7.000 - Psu0 Fan Status Monitor Test ........................... PASS
diagS6000IsPsuGood[954]: ERROR: Psu:1, Power supply is not present.
Test 7.001 - Psu1 Fan Status Monitor Test ........................... NOT
PRESENT
Test 7 - Psu Fan Status Monitor Test ................................ NOT
PRESENT
Test 8.000 - Psu0 Fan AirFlow Type Test ............................. PASS
```

```
diagS6000IsPsuGood[954]: ERROR: Psu:1, Power supply is not present.
Test 8.001 - Psu1 Fan AirFlow Type Test .............................NOT PRESENT
Test 8 - Psu Fan AirFlow Type Test ..................................NOT PRESENT
Test 9 - Power Rail Status Test ..................................... PASS
Test 10.000 - FanTray0 Presence Test ................................ PASS
Test 10.001 - FanTray1 Presence Test ................................ PASS
diagS6000FanTrayPresenceTest[285]: ERROR: FanTray:2 is not present.
Test 10.002 - FanTray2 Presence Test ................................NOT PRESENT
Test 10 - FanTray Presence Test .....................................NOT PRESENT
Test 11.000 - FanTray0 Status Monitor Test .......................... PASS
Test 11.001 - FanTray1 Status Monitor Test .......................... PASS
diagS6000FanStatusMonitorTest[328]: ERROR: FanTray:2 is failed or not present.
Test 11.002 - FanTray2 Status Monitor Test ..........................NOT PRESENT
Test 11 - FanTray Status Monitor Test ...............................NOT PRESENT
Test 12.000 - FanTray0 AirFlow Type Test ............................ PASS
Test 12.001 - FanTray1 AirFlow Type Test ............................ PASS
diagS6000IsFanTrayPresent[442]: ERROR: FanTray:2 is not present.
Test 12.002 - FanTray2 AirFlow Type Test ............................NOT PRESENT
Test 12 - FanTray AirFlow Type Test .................................NOT PRESENT
Test 13.000 - I2c Access Test ....................................... PASS
Test 13.001 - I2c Access Test ....................................... PASS
Test 13.002 - I2c Access Test ....................................... PASS
Test 13.003 - I2c Access Test ....................................... PASS
Test 13.004 - I2c Access Test ....................................... PASS
Test 13.005 - I2c Access Test ....................................... PASS
Test 13.006 - I2c Access Test ....................................... PASS
Test 13.007 - I2c Access Test ....................................... PASS
Test 13.008 - I2c Access Test ....................................... PASS
Test 13.009 - I2c Access Test ....................................... PASS
Test 13.010 - I2c Access Test ....................................... PASS
Test 13.011 - I2c Access Test ....................................... PASS
f10DiagI2cRead[203]: ERROR: i2c_read failed, dev=/dev/i2c2, i2c addr:0x51
offset:0 reglen:0x1 buflen:0x1, flags: 0, rv:-1
diagS6000I2cAccessTest[2606]: ERROR: FanTray2 Eeprom Device --->  ABSENT
Test 13.012 - I2c Access Test ......................................  NOT
PRESENT
Test 13.013 - I2c Access Test ....................................... PASS
Test 13.014 - I2c Access Test ....................................... PASS
Test 13.015 - I2c Access Test ....................................... PASS
Test 13.016 - I2c Access Test ....................................... PASS
diagS6000IsPsuGood[954]: ERROR: Psu:1, Power supply is not present.
diagS6000I2cAccessTest[2722]: ERROR: Psu1 Eeprom Device --->  ABSENT
Test 13.017 - I2c Access Test ....................................... NOT
PRESENT
diagS6000I2cAccessTest[2728]: ERROR: Psu1 PMBus Device --->  ABSENT
Test 13.018 - I2c Access Test ....................................... NOT
PRESENT
Test 13.019 - I2c Access Test .......................................
PASS

     I2C Devices Scanned   - 20
I2C Device PASS Count - 17
I2C Device FAIL Count - 3
Test 13 - I2c Access Test ........................................... FAIL
Test 14 - RTC Presence Test ......................................... PASS
Test 15 - CPU Sdram Presence Test ................................... PASS
Test 16 -  CPU Sdram Size Test ...................................... PASS
Test 17.000 - System Cpld Access Test ............................... PASS
Test 17.001 - Master Cpld Access Test ............................... PASS
Test 17.002 - Slave Cpld Access Test ................................ PASS
Test 17 - Cpld Access Test .......................................... PASS
Test 18 - CFast Presence Test ....................................... PASS
Test 19 - Management Phy Presence Test .............................. PASS
```

Configuring 104 10GbE Ports on the S6000 Platform

The capability to configure up to 104 10-Gigabit Ethernet ports is supported on the S6000 platform.

Starting with Dell Networking OS version 9.4(0.0), you can configure a maximum of 104 10G ports on an S6000 switch.

The network processing unit (NPU) in an S6000 Switch contains two port pipes X and Y, with each pipe supporting a maximum of 52 ports. Therefore, S6000 can support a maximum of 104 ports .

X pipe port mapping: from 0 – 60

Y pipe port mapping: from 64 - 124

When you configure twenty–six 40G ports for Fanout mode, a total of 104 10G ports is available. Because this number is the maximum limit supported on the device, the remaining six 40G ports are disabled.

Configuring Fanout Mode

Keep the following in mind when you specify fanout on S6000 switch ports:

- When you configure the fanout feature, if the maximum limitation of port pipes (52 ports) is reached, the remaining active 40G ports in the port pipe are disabled. In this case, the CLI displays a warning message before the ports are disabled.
- Disabled ports are not available for configuration, are not displayed in the `show` command output, and are not available for SNMP settings. When you reach the maximum limitation of port pipes, a warning message displays.
- The following table describes the fanout feature on ports and the number of available ports per port pipe:

Table 75. Fanout feature on ports

Configuration	Available Ports (Per Pipe)
Each port Pipe: Any 12 Ports fanout.	12 ports fanout (4 * 10G) & 4 40G only ports
Each port Pipe: Any 13 Ports fanout.	13 ports fanout (4 * 10G) & 3 ports disabled

- If you configure 12 ports as fanout in each pipe, the system determines the ports that are not defined for fanout and configures them as 40G ports. Entries in the port table are revised accordingly. If you configure 13 ports as fanout in each pipe, the system identifies the three ports in each pipe that are not configured in Fanout mode and deletes those entries from the port table. This removal of entries ensures that ports are not created for such pipes.
- Entity MIB that uses static port configuration supports all of the ports, including those ports that are configured for fanout.
- When you configure 26 ports as fanout, six ports are disabled. For disabled ports, interfaces are not created.
- When you insert an optic into one of the disabled ports, a system logging message generates indicating that a 40GE QSFP optic has been inserted into a disabled port.

- You can configure a disabled port as a fanout port when the maximum limit is not exceeded. This behavior reduces the number of reloads required to configure a disabled port in Quad mode to Non-quad mode.
- When you configure 13 ports as fanout, a warning message is displayed stating that the maximum number of ports that you can configure in Quad mode has been reached. The extra ports are disabled the next time you reload the system.
- When you attempt to configure fanout on the ports where fanout feature is not supported, a warning message displays stating that Quad mode ports limit has been reached and that you cannot configure additional ports.
- For the`show hardware` command, the valid range for the port-range parameter is from 0-127.
- Internal ports mapped to global ports may vary depending on the configured attributes. The `show hardware` command output using these internal ports is realigned according to the configured settings.
- To view the disabled ports, you can use the`disabled-ports` keyword with the `show system stack-unit 0` command.
- Before you configure the last available port in Quad mode, you are prompted to confirm whether you want to configure the port Fanout mode, as shown in the following example.

```
Dell(conf)#stack-unit 0 port 0 portmode quad
Maximum number of ports that can be made Quad mode in the range <0-63> is
configured. Ports 12,16,60, will be disabled on reload.
Do you wish to continue? [confirm yes/no]:yes
Please save and reset unit 0 for the changes to take effect.
```

The following figure shows 26 ports configured in Quad mode. Ports X(52, 56, 60) and Y(64, 68, 72) are disabled (shown in red).

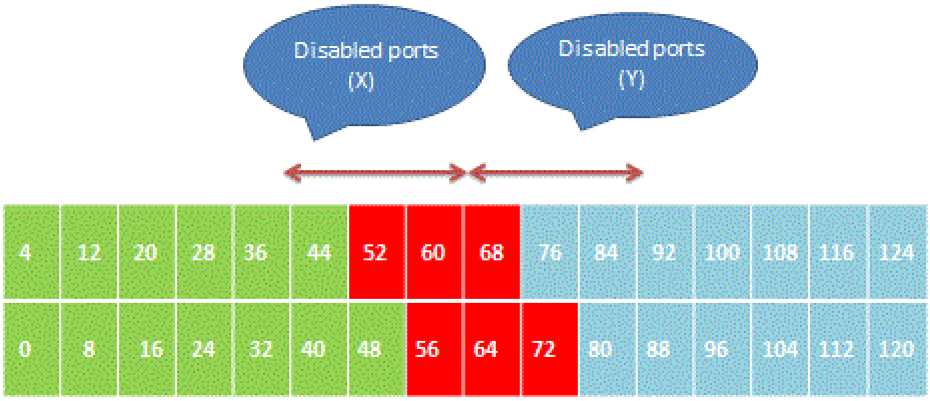


Figure 129. Port configured in Quad mode.

Backward Compatibility With Enhanced Fanout Configuration

When you configure fanout on the ports in which fanout is not supported in releases of Dell Networking OS earlier than Release 9.4(0.0), downgrade to an earlier release might cause an unexpected behavior.

This phenomenon occurs because in Release 9.4(0.0), static configuration for the restricted ports is maintained. Validation of the restricted ports is performed at the CLI level and not during bootup. Therefore, increased fanout configuration on ports is not backward-compatible.

If you want to perform a downgrade from Release 9.3(0.0) to 9.0(2.0), Dell Networking recommends removing the fanout configuration on the ports that were previously not enabled for fanout. However, if you want to perform a downgrade from Release 9.4(0.0) to Release 9.3(0.0), you must not remove fanout on restricted ports. Fanout configuration in NVRAM is not reset for ports 4, 12, 20, 28, 100, 108, 116, and 124, the ports where fanout configuration is not allowed. The ports are reset and considered as 40G ports.

Configuring Fanout on Disabled Ports

When you configure fanout on the maximum permissible number of 40G ports, the remaining ports are disabled on the port pipe. If you want to configure fanout on the disabled ports, you must first enable the disabled ports by removing the fanout setting from the active ports and then reconfigure fanout on the desired port. This operation requires you to reboot the system twice: once to enable the port and another time to configure fanout because the fanout setting becomes effective only after a reboot. You can avoid this situation by configuring the disabled ports to be in Quad mode if you don not exceed the maximum limit on fanout ports.

When you use the `show hardware stack-unit 0 drops unit 0` command, you do not need to know which are 40G ports because the limitation in Release 9.3(0.0) is no longer valid. The following sample output displays the user port and system port mapping if you configure all the ports as 40G.

```
Dell#show hardware stack-unit 0 drops unit 0

UserPort PortNumber Ingress Drops  IngMac Drops  Total Mmu Drops  EgMac Drops
Egress Drops
 0        1           0              0              0               0
0
 4        5           0              0              0               0
0
 8        9           0              0              0               0
0
 12       13          0              0              0               0
0
 16       17          0              0              0               0
0
 20       21          0              0              0               0
0
 24       25          0              0              0               0
0
 28       29          0              0              0               0
0
 32       33          0              0              0               0
0
 36       37          0              0              0               0
0
 40       41          0              0              0               0
0
 44       45          0              0              0               0
0
 48       49          0              0              0               0
0
 52       50          0              0              0               0
0
 56       51          0              0              0               0
```

```
0
 60       52          0            0              0              0
0
 64       53          0            0              0              0
0
 68       57          0            0              0              0
0
 72       61          0            0              0              0
0
 76       65          0            0              0              0
0
 80       69          0            0              0              0
0
 84       73          0            0              0              0
0
 88       77          0            0              0              0
0
 92       81          0            0              0              0
0
 96       85          0            0              0              0
0
 100      89          0            0              0              0
0
 104      93          0            0              0              0
0
 108      97          0            0              0              0
0
 112      101         0            0              0              0
0
 116      102         0            0              0              0
0
 12       103         0            0              0              0
0
 1        104         0            0              0              0
0
```

Trace Logs

In addition to the syslog buffer, Dell Networking OS buffers trace messages which are continuously written by various Dell Networking OS software tasks to report hardware and software events and status information.

Each trace message provides the date, time, and name of the Dell Networking OS process. All messages are stored in a ring buffer. You can save the messages to a file either manually or automatically after failover.

Auto Save on Crash or Rollover

Exception information for MASTER or standby units is stored in the *flash:/TRACE_LOG_DIR* directory. This directory contains files that save trace information when there has been a task crash or timeout.

- On a MASTER unit, you can reach the *TRACE_LOG_DIR* files by FTP or by using the `show file` command from the *flash://TRACE_LOG_DIR* directory.
- On a Standby unit, you can reach the *TRACE_LOG_DIR* files only by using the `show file` command from the *flash://TRACE_LOG_DIR* directory.

NOTE: Non-management member units do not support this functionality.

Hardware Watchdog Timer

The `hardware watchdog` command automatically reboots an Dell Networking OS switch/router with a single RPM that is unresponsive.

This is a last resort mechanism intended to prevent a manual power cycle.

Enabling Environmental Monitoring

The components use environmental monitoring hardware to detect transmit power readings, receive power readings, and temperature updates.
To receive periodic power updates, you must enable the following command.

- Enable environmental monitoring.

  `enable optic-info-update interval`

Example of the `show interfaces transceiver` Command

```
Dell#show interfaces transceiver

-- RPM Environment Status --
Slot Status     Temp  Voltage
------------------------------
0    active     33C   ok
1    not present
```

Recognize an Overtemperature Condition

An overtemperature condition occurs, for one of two reasons: the card genuinely is too hot or a sensor has malfunctioned.

Inspect cards adjacent to the one reporting the condition to discover the cause.

- If directly adjacent cards are not normal temperature, suspect a genuine overheating condition.
- If directly adjacent cards are normal temperature, suspect a faulty sensor.

When the system detects a genuine over-temperature condition, it powers off the card. To recognize this condition, look for the following system messages:

```
CHMGR-2-MAJOR_TEMP: Major alarm: chassis temperature high (temperature reaches
or exceeds threshold of
[value]C)
CHMGR-2-TEMP_SHUTDOWN_WARN: WARNING! temperature is [value]C; approaching
shutdown threshold of [value]C
```

To view the programmed alarm thresholds levels, including the shutdown value, use the `show alarms threshold` command.

Example of the `show alarms threshold` Command

```
Dell#show alarms threshold

-- Temperature Limits (deg C) --
------------------------------------------
         Minor Minor Off Major Major Off Shutdown
Linecard 75    70        80    77        85
```

```
RPM       65    60         75    70         80
Dell#
```

Troubleshoot an Over-temperature Condition

To troubleshoot an over-temperature condition, use the following information.

1. Use the `show environment` commands to monitor the temperature levels.
2. Check air flow through the system. Ensure that the air ducts are clean and that all fans are working correctly.
3. After the software has determined that the temperature levels are within normal limits, you can re-power the card safely. To bring back the line card online, use the `power-on` command in EXEC mode.

In addition, to control airflow for adequate system cooling, Dell Networking requires that you install blanks in all slots without a line card.

NOTE: Exercise care when removing a card; if it has exceeded the major or shutdown thresholds, the card could be hot to the touch.

Recognize an Under-Voltage Condition

If the system detects an under-voltage condition, it sends an alarm.

To recognize this condition, look for the following system message: `%CHMGR-1-CARD_SHUTDOWN: Major alarm: Line card 2 down - auto-shutdown due to under voltage.`

This message indicates that the specified card is not receiving enough power. In response, the system first shuts down Power over Ethernet (PoE). If the under-voltage condition persists, line cards are shut down, then the RPMs.

Troubleshoot an Under-Voltage Condition

To troubleshoot an under-voltage condition, check that the correct number of power supplies are installed and their Status light emitting diodes (LEDs) are lit.

The following table lists information for SNMP traps and OIDs on S-Series environmental monitoring hardware and hardware components.

Table 76. SNMP Traps and OIDs

OID String	OID Name	Description
Receiving Power		
.1.3.6.1.4.1.6027.3.10.1.2.5.1.6	chSysPortXfpRecvPower	OID displays the receiving power of the connected optics.
Transmitting power		
.1.3.6.1.4.1.6027.3.10.1.2.5.1.8	chSysPortXfpTxPower	OID displays the transmitting power of the connected optics.
Temperature		
.1.3.6.1.4.1.6027.3.10.1.2.5.1.7	chSysPortXfpRecvTemp	OID displays the temperature of the connected optics.

OID String	OID Name	Description
		NOTE: These OIDs only generate if you enable the `enable optic-info-update-interval is enabled` command.
Hardware MIB Buffer Statistics		
.1.3.6.1.4.1.6027.3.16.1.1.4	fpPacketBufferTable	View the modular packet buffers details per stack unit and the mode of allocation.
.1.3.6.1.4.1.6027.3.16.1.1.5	fpStatsPerPortTable	View the forwarding plane statistics containing the packet buffer usage per port per stack unit.
.1.3.6.1.4.1.6027.3.16.1.1.6	fpStatsPerCOSTable	View the forwarding plane statistics containing the packet buffer statistics per COS per port.

Buffer Tuning

Buffer tuning allows you to modify the way your switch allocates buffers from its available memory and helps prevent packet drops during a temporary burst of traffic.

The S-Series application-specific integrated circuit (ASICs) implement the key functions of queuing, feature lookups, and forwarding lookups in hardware.

Forwarding processor (FP) ASICs provide Ethernet MAC functions, queueing, and buffering, as well as store feature and forwarding tables for hardware-based lookup and forwarding decisions. 1G and 10G interfaces use different FPs.

The following table describes the type and number of ASICs per platform.

Table 77. ASICs by Platform

Hardware	FP	CSF
S50N, S50V	2	0
S25V, S25P, S25N	1	0

As shown in the following example, you can tune buffers at three locations.

1. CSF — Output queues going from the CSF.
2. FP Uplink — Output queues going from the FP to the CSF IDP links.
3. Front-End Link — Output queues going from the FP to the front-end PHY.

All ports support eight queues, four for data traffic and four for control traffic. All eight queues are tunable.

Physical memory is organized into cells of 128 bytes. The cells are organized into two buffer pools — the dedicated buffer and the dynamic buffer.

- **Dedicated buffer** — this pool is reserved memory that other interfaces cannot use on the same ASIC or by other queues on the same interface. This buffer is always allocated, and no dynamic re-carving takes place based on changes in interface status. Dedicated buffers introduce a trade-off. They provide each interface with a guaranteed minimum buffer to prevent an overused and congested interface from starving all other interfaces. However, this minimum guarantee means that the buffer manager does not reallocate the buffer to an adjacent congested interface, which means that in some cases, memory is under-used.
- **Dynamic buffer** — this pool is shared memory that is allocated as needed, up to a configured limit. Using dynamic buffers provides the benefit of statistical buffer sharing. An interface requests dynamic buffers when its dedicated buffer pool is exhausted. The buffer manager grants the request based on three conditions:
  - The number of used and available dynamic buffers.
  - The maximum number of cells that an interface can occupy.
  - Available packet pointers (2k per interface). Each packet is managed in the buffer using a unique packet pointer. Thus, each interface can manage up to 2k packets.

You can configure dynamic buffers per port on both 1G and 10G FPs and per queue on CSFs. By default, the FP dynamic buffer allocation is 10 times oversubscribed. For the 48-port 1G card:

- Dynamic Pool= Total Available Pool(16384 cells) — Total Dedicated Pool = 5904 cells
- Oversubscription ratio = 10
- Dynamic Cell Limit Per port = 59040/29 = 2036 cells

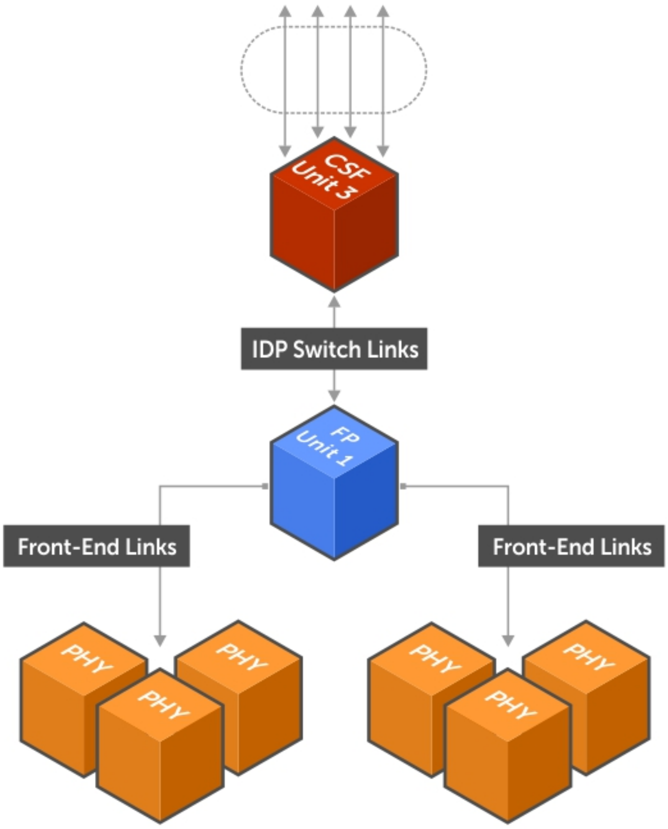


Figure 130. Buffer Tuning Points

Deciding to Tune Buffers

Dell Networking recommends exercising caution when configuring any non-default buffer settings, as tuning can significantly affect system performance. The default values work for most cases.
As a guideline, consider tuning buffers if traffic is bursty (and coming from several interfaces). In this case:

- Reduce the dedicated buffer on all queues/interfaces.
- Increase the dynamic buffer on all interfaces.
- Increase the cell pointers on a queue that you are expecting will receive the largest number of packets.

To define, change, and apply buffers, use the following commands.

- Define a buffer profile for the FP queues.

  CONFIGURATION mode

  ```
  buffer-profile fp fsqueue
  ```
- Define a buffer profile for the CSF queues.

  CONFIGURATION mode

  ```
  buffer-profile csf csqueue
  ```
- Change the dedicated buffers on a physical 1G interface.

BUFFER PROFILE mode

```
buffer dedicated
```

- Change the maximum number of dynamic buffers an interface can request.

  BUFFER PROFILE mode

  ```
  buffer dynamic
  ```

- Change the number of packet-pointers per queue.

  BUFFER PROFILE mode

  ```
  buffer packet-pointers
  ```

- Apply the buffer profile to a line card.

  CONFIGURATION mode

  ```
  buffer fp-uplink linecard
  ```

- Apply the buffer profile to a CSF to FP link.

  CONFIGURATION mode

  ```
  buffer csf linecard
  ```

Examples of Viewing Buffer Information

Dell Networking OS Behavior: If you attempt to apply a buffer profile to a non-existent port-pipe, Dell Networking OS displays the following message: `%DIFFSERV-2-DSA_BUFF_CARVING_INVALID_PORT_SET: Invalid FP port-set 2 for linecard 2. Valid range of port-set is <0-1>`. However, the configuration still appears in the running-config.

Configuration changes take effect immediately and appear in the running configuration. Because under normal conditions all ports do not require the maximum allocation, the configured dynamic allocations can exceed the actual amount of available memory; this allocation is called oversubscription. If you choose to oversubscribe the dynamic allocation, a burst of traffic on one interface might prevent other interfaces from receiving the configured dynamic allocation, which causes packet loss.

You cannot allocate more than the available memory for the dedicated buffers. If the system determines that the sum of the configured dedicated buffers allocated to the queues is more than the total available memory, the configuration is rejected, returning a syslog message similar to the following: `00:04:20: %S50N:0 %DIFFSERV-2-DSA_DEVICE_BUFFER_UNAVAILABLE: Unable to allocate dedicated buffers for stack-unit 0, port pipe 0, egress port 25 due to unavailability of cells`.

Dell Networking OS Behavior: When you remove a buffer-profile using the `no buffer-profile [fp | csf]` command from CONFIGURATION mode, the buffer-profile name still appears in the output of the `show buffer-profile [detail | summary]` command. After a line card reset, the buffer profile correctly returns to the default values, but the profile name remains. Remove it from the `show buffer-profile [detail | summary]` command output by entering `no buffer [fp-uplink |csf] linecard port-set buffer-policy` from CONFIGURATION mode and `no buffer-policy` from INTERFACE mode.

To display the allocations for any buffer profile, use the `show` commands.

To display the default buffer profile, use the `show buffer-profile {summary | detail}` command from EXEC Privilege mode.

The following example shows viewing the default buffer profile.

```
Dell#show buffer-profile detail interface gigabitethernet 0/1
Interface Gi 0/1
Buffer-profile -
Dynamic buffer 194.88 (Kilobytes)
Queue# Dedicated Buffer  Buffer Packets
       (Kilobytes)
0      2.50              256
1      2.50              256
2      2.50              256
3      2.50              256
4      9.38              256
5      9.38              256
6      9.38              256
7      9.38              256
```

The following example shows viewing the buffer profile allocations.

```
Dell#show running-config interface tengigabitethernet 2/0 !
interface TenGigabitEthernet 2/0
no ip address
mtu 9252
switchport
no shutdown
buffer-policy myfsbufferprofile
```

The following example shows viewing the default buffer profile on an interface.

```
Dell#sho buffer-profile detail int gi 0/10
Interface Gi 0/10
Buffer-profile fsqueue-fp
Dynamic buffer 1256.00 (Kilobytes)
Queue# Dedicated Buffer  Buffer Packets
       Kilobytes)
0      3.00              256
1      3.00              256
2      3.00              256
3      3.00              256
4      3.00              256
5      3.00              256
6      3.00              256
7      3.00              256
```

The following example shows viewing the default buffer profile on a linecard.

```
Dell#sho buffer-profile detail fp-uplink stack-unit 0 port-set 0
Linecard 0 Port-set 0
Buffer-profile fsqueue-hig
Dynamic Buffer 1256.00 (Kilobytes)
Queue# Dedicated Buffer Buffer Packets
       (Kilobytes)
0      3.00             256
1      3.00             256
2      3.00             256
3      3.00             256
4      3.00             256
5      3.00             256
6      3.00             256
7      3.00             256
```

Using a Pre-Defined Buffer Profile

Dell Networking OS provides two pre-defined buffer profiles, one for single-queue (for example, non-quality-of-service [QoS]) applications, and one for four-queue (for example, QoS) applications.
You must reload the system for the global buffer profile to take effect, a message similar to the following displays: `% Info: For the global pre-defined buffer profile to take effect, please save the config and reload the system..`

Dell Networking OS Behavior: After you configure buffer-profile global 1Q, the message displays during every bootup. Only one reboot is required for the configuration to take effect; afterward you may ignore this bootup message.

Dell Networking OS Behavior: If you configure 1Q, save the running-config to the startup-config, and then delete the startup-config and reload the chassis. The only way to return to the default buffer profile is to remove the 1Q profile configured and then reload the chassis.

If you have already applied a custom buffer profile on an interface, the `buffer-profile global` command fails and a message similar to the following displays: `% Error: User-defined buffer profile already applied. Failed to apply global pre-defined buffer profile. Please remove all user-defined buffer profiles.`

Similarly, when you configure `buffer-profile global`, you cannot not apply a buffer profile on any single interface. A message similar to the following displays: `% Error: Global pre-defined buffer profile already applied. Failed to apply user-defined buffer profile on interface Gi 0/1. Please remove global pre-defined buffer profile.`

To apply a predefined buffer profile, use the following command.

- Apply one of the pre-defined buffer profiles for all port pipes in the system.
  CONFIGURATION mode

  ```
  buffer-profile global [1Q|4Q]
  ```

If the default buffer profile `dynamic` is active, Dell Networking OS displays an error message instructing you to remove the default configuration using the `no buffer-profile global` command.

Sample Buffer Profile Configuration

The two general types of network environments are sustained data transfers and voice/data.

Dell Networking recommends a single-queue approach for data transfers.
Example of a Single Queue Application for S50N with Default Packet Pointers

```
!
buffer-profile fp fsqueue-fp
buffer dedicated queue0 3 queue1 3 queue2 3 queue3 3 queue4 3 queue5 3 queue6 3 
queue7 3
buffer dynamic 1256
!
  buffer-profile fp fsqueue-hig
  buffer dedicated queue0 3 queue1 3 queue2 3 queue3 3 queue4 3 queue5 3 queue6 
3 queue7 3
  buffer dynamic 1256

!
buffer fp-uplink stack-unit 0 port-set 0 buffer-policy fsqueue-hig
```

```
buffer fp-uplink stack-unit 0 port-set 1 buffer-policy fsqueue-hig
!
Interface range gi 0/1 - 48
buffer-policy fsqueue-fp

Dell#sho run int gi 0/10
!
interface GigabitEthernet 0/10
  no ip address
```

Troubleshooting Packet Loss

The `show hardware stack-unit` command is intended primarily to troubleshoot packet loss. To troubleshoot packet loss, use the following commands.

- `show hardware stack-unit cpu data-plane statistics`
- `show hardware stack-unit cpu party-bus statistics`
- `show hardware stack-unit` *`0-11`* `drops unit` *`0-1`* `port` *`0-63`*
- `show hardware stack-unit` *`0-11`* `stack-port` *`48-51`*
- `show hardware stack-unit` *`0-11`* `unit` *`0-1`* `{counters | details | port-stats [detail] | register | execute-shell-cmd | ipmc-replication | table-dump}`
- `show hardware {layer2| layer3} {e.g. acl |in acl} stack-unit` *`0-11`* `port-set` *`0-1`*
- `show hardware layer3 qos stack-unit` *`0-`*`7 port-set` *`0-1`*
- `show hardware ipv6 {e.g.-acl |in-acl} stack-unit` *`0-11`* `port-set` *`0-1`*
- `show hardware system-flow layer2 stack-unit` *`0-11`* `port-set` *`0-1`* `[counters]`
- `clear hardware stack-unit` *`0-11`* `counters`
- `clear hardware stack-unit` *`0-11`* `unit` *`0-1`* `counters`
- `clear hardware stack-unit` *`0-11`* `cpu data-plane statistics`
- `clear hardware stack-unit` *`0-11`* `cpu party-bus statistics`
- `clear hardware stack-unit` *`0-11`* `stack-port` *`48-51`*

Displaying Drop Counters

To display drop counters, use the following commands.

- Identify which stack unit, port pipe, and port is experiencing internal drops.

  `show hardware stack-unit` *`0-11`* `drops [unit 0 [`*`port 0-63`*`]]`
- Display drop counters.

  `show hardware stack-unit drops unit port`

Example of the `show hardware stack-unit` Command to View Drop Counters Statistics

```
Dell#show hardware stack-unit 0 drops
UNIT No: 0
Total Ingress Drops :0
Total IngMac Drops :0
Total Mmu Drops :0
Total EgMac Drops :0
Total Egress Drops :0
UNIT No: 1
Total Ingress Drops :0
Total IngMac Drops :0
Total Mmu Drops :0
Total EgMac Drops :0
Total Egress Drops :0
```

```
Dell#show hardware stack-unit 0 drops unit 0
Port# :Ingress Drops :IngMac Drops :Total Mmu Drops :EgMac Drops :Egress
Drops
1 0 0 0 0 0
2 0 0 0 0 0
3 0 0 0 0 0
4 0 0 0 0 0
5 0 0 0 0 0
6 0 0 0 0 0
7 0 0 0 0 0
8 0 0 0 0 0


Dell#show hardware stack-unit 0 drops unit 0 port 1
--- Ingress Drops ---
Ingress Drops                : 30
IBP CBP Full Drops           : 0
PortSTPnotFwd Drops          : 0
IPv4 L3 Discards             : 0
Policy Discards              : 0
Packets dropped by FP        : 14
(L2+L3) Drops                : 0
Port bitmap zero Drops       : 16
Rx VLAN Drops                : 0

--- Ingress MAC counters---
Ingress FCSDrops             : 0
Ingress MTUExceeds           : 0

--- MMU Drops ---
HOL DROPS                    : 0
TxPurge CellErr              : 0
Aged Drops                   : 0

--- Egress MAC counters---
Egress FCS Drops             : 0

--- Egress FORWARD PROCESSOR Drops ---
IPv4 L3UC Aged & Drops       : 0
TTL Threshold Drops          : 0
INVALID VLAN CNTR Drops      : 0
L2MC Drops                   : 0
PKT Drops of ANY Conditions  : 0
Hg MacUnderflow              : 0
TX Err PKT Counter           : 0
```

Dataplane Statistics

The `show hardware stack-unit cpu data-plane statistics` command provides insight into the packet types coming to the CPU. The `show hardware stack-unit cpu party-bus statistics` command displays input and output statistics on the party bus, which carries inter-process communication traffic between CPUs.

The command output in the following example has been augmented, providing detailed RX/ TX packet statistics on a per-queue basis. The objective is to see whether CPU-bound traffic is internal (so-called *party bus* or IPC traffic) or network control traffic, which the CPU must process.

Example of Viewing Dataplane Statistics

```
Dell#show hardware stack-unit 2 cpu data-plane statistics

bc pci driver statistics for device:
  rxHandle        :0
  noMhdr          :0
  noMbuf          :0
  noClus          :0
  recvd           :0
  dropped         :0
  recvToNet       :0
  rxError         :0
  rxDatapathErr   :0
  rxPkt(COS0)     :0
  rxPkt(COS1)     :0
  rxPkt(COS2)     :0
  rxPkt(COS3)     :0
  rxPkt(COS4)     :0
  rxPkt(COS5)     :0
  rxPkt(COS6)     :0
  rxPkt(COS7)     :0
  rxPkt(UNIT0)    :0
  rxPkt(UNIT1)    :0
  rxPkt(UNIT2)    :0
  rxPkt(UNIT3)    :0
  transmitted     :0
  txRequested     :0
  noTxDesc        :0
  txError         :0
  txReqTooLarge   :0
  txInternalError :0
  txDatapathErr   :0
  txPkt(COS0)     :0
  txPkt(COS1)     :0
  txPkt(COS2)     :0
  txPkt(COS3)     :0
  txPkt(COS4)     :0
  txPkt(COS5)     :0
  txPkt(COS6)     :0
  txPkt(COS7)     :0
  txPkt(UNIT0)    :0
```

Example of Viewing Party Bus Statistics

```
Dell#sh hardware stack-unit 2 cpu party-bus statistics
Input Statistics:
  27550 packets, 2559298 bytes
  0 dropped, 0 errors
Output Statistics:
  1649566 packets, 1935316203 bytes
  0 errors
```

Display Stack Port Statistics

The `show hardware stack-unit stack-port` command displays input and output statistics for a stack-port interface.

Example of Viewing Stack Unit Statistics

```
Dell#show hardware stack-unit 2 stack-port 49
Input Statistics:
  27629 packets, 3411731 bytes
  0 64-byte pkts, 27271 over 64-byte pkts, 207 over 127-byte pkts
  17 over 255-byte pkts, 56 over 511-byte pkts, 78 over 1023-byte pkts
  0 Multicasts, 5 Broadcasts
  0 runts, 0 giants, 0 throttles
  0 CRC, 0 overrun, 0 discarded
Output Statistics:
  1649714 packets, 1948622676 bytes, 0 underruns
  0 64-byte pkts, 27234 over 64-byte pkts, 107970 over 127-byte pkts
  34 over 255-byte pkts, 504838 over 511-byte pkts, 1009638 over 1023-byte pkts
  0 Multicasts, 0 Broadcasts, 1649714 Unicasts
  0 throttles, 0 discarded, 0 collisions
Rate info (interval 45 seconds):
  Input 00.00 Mbits/sec,    2 packets/sec, 0.00% of line-rate
  Output 00.06 Mbits/sec,   8 packets/sec, 0.00% of line-rate
Dell#
```

Display Stack Member Counters

The `show hardware stack-unit` *`0–11`* `{counters | details | port-stats [detail] | register}` command displays internal receive and transmit statistics, based on the selected command option.

The following example is a sample of the output for the `counters` option.

Example of Displaying Stack Unit Counters

```
RIPC4.ge0   :   1,202     +1,202
RUC.ge0     :   1,224     +1,217
RDBGC0.ge0  :   34        +24
RDBGC1.ge0  :   366       +235
RDBGC5.ge0  :   16        +12
RDBGC7.ge0  :   18        +12
GR64.ge0    :   5,176     +24
GR127.ge0   :   1,566     +1,433
GR255.ge0   :   4         +4
GRPKT.ge0   :   1,602     +1,461
GRBYT.ge0   :   117,600   +106,202
GRMCA.ge0   :   366       +235
GRBCA.ge0   :   12        +9
GT64.ge0    :   4         +3
GT127.ge0   :   964       +964
GT255.ge0   :   4         +4
GT511.ge0   :   1         +1
GTPKT.ge0   :   973       +972
GTBCA.ge0   :   1         +1
GTBYT.ge0   :   71,531    +71,467
RUC.cpu0    :   972       +971
TDBGC6.cpu0 :   1,584     +1,449=
```

Enabling Application Core Dumps

Application core dumps are disabled by default.
A core dump file can be very large. Due to memory requirements the file can only be sent directly to an FTP server; it is not stored on the local flash.

To enable full application core dumps, use the following command.

- Enable RPM core dumps and specify the Shutdown mode.

  CONFIGURATION mode

  ```
  logging coredump server
  ```

To undo this command, use the `no logging coredump server` command.

Mini Core Dumps

Dell Networking OS supports mini core dumps on the application and kernel crashes. The mini core dump applies to Master, Standby, and Member units.

Application and kernel mini core dumps are always enabled. The mini core dumps contain the stack space and some other minimal information that you can use to debug a crash. These files are small files and are written into flash until space is exhausted. When the flash is full, the write process is stopped.

A mini core dump contains critical information in the event of a crash. Mini core dump files are located in *flash:/ (root dir)*. The application mini core filename format is *f10StkUnit<Stack_unit_no>.<Application name>.acore.mini.txt*. The kernel mini core filename format is *f10StkUnit<Stack_unit_no>.kcore.mini.txt*. The following are sample filenames.

When a member or standby unit crashes, the mini core file gets uploaded to master unit. When the master unit crashes, the mini core file is uploaded to new master.

The panic string contains key information regarding the crash. Several panic string types exist, and they are displayed in regular English text to enable easier understanding of the crash cause.

Example of Application Mini Core Dump Listings

```
Dell#dir
Directory of flash:

1  drw-  16384 Jan 01 1980 00:00:00 +00:00 .
2  drwx  1536  Sep 03 2009 16:51:02 +00:00 ..
3  drw-  512   Aug 07 2009 13:05:58 +00:00 TRACE_LOG_DIR
4  d---  512   Aug 07 2009 13:06:00 +00:00 ADMIN_DIR
5  -rw-  8693  Sep 03 2009 16:50:56 +00:00 startup-config
6  -rw-  8693  Sep 03 2009 16:44:22 +00:00 startup-config.bak
7  -rw-  156   Aug 28 2009 16:16:10 +00:00 f10StkUnit0.mrtm.acore.mini.txt
8  -rw-  156   Aug 28 2009 17:17:24 +00:00 f10StkUnit0.vrrp.acore.mini.txt
9  -rw-  156   Aug 28 2009 18:25:18 +00:00 f10StkUnit0.sysd.acore.mini.txt
10 -rw-  156   Aug 28 2009 19:07:36 +00:00 f10StkUnit0.frrp.acore.mini.txt
11 -rw-  156   Aug 31 2009 16:18:50 +00:00 f10StkUnit2.sysd.acore.mini.txt
12 -rw-  156   Aug 29 2009 14:28:34 +00:00 f10StkUnit0.ipm1.acore.mini.txt
13 -rw-  156   Aug 31 2009 16:14:56 +00:00 f10StkUnit0.acl.acore.mini.txt

flash: 3104256 bytes total (2959872 bytes free)
Dell#
```

Example of a Mini Core Text File

```
VALID MAGIC
------------------PANIC STRING -----------------
panic string is :<null>
----------------STACK TRACE START---------------
0035d60c <f10_save_mmu+0x120>:
00274f8c <panic+0x144>:
0024e2b0 <db_fncall+0x134>:
0024dee8 <db_command+0x258>:
0024d9c4 <db_command_loop+0xc4>:
002522b0 <db_trap+0x158>:
0026a8d0 <mi_switch+0x1b0>:
0026a00c <bpendtsleep>:
-----------------STACK TRACE END----------------
```

```
--------------------FREE MEMORY---------------
uvmexp.free = 0x2312
```

Enabling TCP Dumps

A TCP dump captures CPU-bound control plane traffic to improve troubleshooting and system manageability. When you enable TCP dump, it captures all the packets on the local CPU, as specified in the CLI.
You can save the traffic capture files to flash, FTP, SCP, or TFTP. The files saved on the flash are located in the *flash://TCP_DUMP_DIR/Tcpdump_<time_stamp_dir>/* directory and labeled *tcpdump_*.pcap*. There can be up to 20 *Tcpdump_<time_stamp_dir>* directories. The 21st file overwrites the oldest saved file. The maximum file size for a TCP dump capture is 1MB. When a file reaches 1MB, a new file is created, up to the specified total number of files.

Maximize the number of packets recorded in a file by specifying the snap-length to capture the file headers only.

The `tcpdump` command has a finite run process. When you enable the `tcpdump` command, it runs until the capture-duration timer and/or the packet-count counter threshold is met. If you do not set a threshold, the system uses a default of a 5 minute capture-duration and/or a single 1k file as the stopping point for the dump.

You can use the capture-duration timer and the packet-count counter at the same time. The TCP dump stops when the first of the thresholds is met. That means that even if the duration timer is 9000 seconds, if the maximum file count parameter is met first, the dumps stop.

To enable a TCP dump, use the following command.

- Enable a TCP dump for CPU bound traffic.
  CONFIGURATION mode

  ```
  tcpdump cp [capture-duration time | filter expression | max-file-count value
  | packet-count value | snap-length value | write-to path]
  ```

57

Standards Compliance

This chapter describes standards compliance for Dell Networking products.

NOTE: Unless noted, when a standard cited here is listed as supported by the Dell Networking Operating System (OS), Dell Networking OS also supports predecessor standards. One way to search for predecessor standards is to use the http://tools.ietf.org/ website. Click "Browse and search IETF documents," enter an RFC number, and inspect the top of the resulting document for obsolescence citations to related RFCs.

IEEE Compliance

The following is a list of IEEE compliance.

802.1AB	LLDP
802.1D	Bridging, STP
802.1p	L2 Prioritization
802.1Q	VLAN Tagging, Double VLAN Tagging, GVRP
802.1s	MSTP
802.1w	RSTP
802.1X	Network Access Control (Port Authentication)
802.3ab	Gigabit Ethernet (1000BASE-T)
802.3ac	Frame Extensions for VLAN Tagging
802.3ad	Link Aggregation with LACP
802.3ae	10 Gigabit Ethernet (10GBASE-W, 10GBASE-X)
802.3af	Power over Ethernet
802.3ak	10 Gigabit Ethernet (10GBASE-CX4)
802.3i	Ethernet (10BASE-T)
802.3u	Fast Ethernet (100BASE-FX, 100BASE-TX)
802.3x	Flow Control
802.3z	Gigabit Ethernet (1000BASE-X)
ANSI/TIA-1057	LLDP-MED
Force10	FRRP (Force10 Redundant Ring Protocol)
Force10	PVST+
SFF-8431	SFP+ Direct Attach Cable (10GSFP+Cu)

MTU	9,252 bytes

RFC and I-D Compliance

Dell Networking OS supports the following standards. The standards are grouped by related protocol. The columns showing support by platform indicate which version of Dell Networking OS first supports the standard.

General Internet Protocols

The following table lists the Dell Networking OS support per platform for general internet protocols.

Table 78. General Internet Protocols

RFC#	Full Name	S-Series
768	User Datagram Protocol	7.6.1
793	Transmission Control Protocol	7.6.1
854	Telnet Protocol Specification	7.6.1
959	File Transfer Protocol (FTP)	7.6.1
1321	The MD5 Message-Digest Algorithm	7.6.1
1350	The TFTP Protocol (Revision 2)	7.6.1
1661	The Point-to-Point Protocol (PPP)	
1989	PPP Link Quality Monitoring	
1990	The PPP Multilink Protocol (MP)	
1994	PPP Challenge Handshake Authentication Protocol (CHAP)	
2460	Internationalization of the File Transfer Protocol	8.3.12.0
2474	Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers	7.7.1
2615	PPP over SONET/SDH	
2698	A Two Rate Three Color Marker	
3164	The BSD syslog Protocol	7.6.1
draft-ietf-bfd -base-03	Bidirectional Forwarding Detection	

General IPv4 Protocols

The following table lists the Dell Networking OS support per platform for general IPv4 protocols.

Table 79. General IPv4 Protocols

RFC#	Full Name	S-Series
791	Internet Protocol	7.6.1
792	Internet Control Message Protocol	7.6.1
826	An Ethernet Address Resolution Protocol	7.6.1
1027	Using ARP to Implement Transparent Subnet Gateways	7.6.1
1035	DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION (client)	7.6.1
1042	A Standard for the Transmission of IP Datagrams over IEEE 802 Networks	7.6.1
1191	Path MTU Discovery	7.6.1
1305	Network Time Protocol (Version 3) Specification, Implementation and Analysis	7.6.1
1519	Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy	7.6.1
1542	Clarifications and Extensions for the Bootstrap Protocol	7.6.1
1812	Requirements for IP Version 4 Routers	7.6.1
2131	Dynamic Host Configuration Protocol	7.6.1
2338	Virtual Router Redundancy Protocol (VRRP)	7.6.1
3021	Using 31-Bit Prefixes on IPv4 Point-to-Point Links	7.7.1
3046	DHCP Relay Agent Information Option	7.8.1
3069	VLAN Aggregation for Efficient IP Address Allocation	7.8.1
3128	Protection Against a Variant of the Tiny Fragment Attack	7.6.1

General IPv6 Protocols

The following table lists the Dell Networking OS support per platform for general IPv6 protocols.

Table 80. General IPv6 Protocols

RFC#	Full Name	S-Series
1886	DNS Extensions to support IP version 6	7.8.1
1981 (Partial)	Path MTU Discovery for IP version 6	7.8.1
2460	Internet Protocol, Version 6 (IPv6) Specification	7.8.1
2462 (Partial)	IPv6 Stateless Address Autoconfiguration	7.8.1
2464	Transmission of IPv6 Packets over Ethernet Networks	7.8.1
2675	IPv6 Jumbograms	7.8.1
2711	IPv6 Router Alert Option	8.3.12.0
3587	IPv6 Global Unicast Address Format	7.8.1
4007	IPv6 Scoped Address Architecture	8.3.12.0
4291	Internet Protocol Version 6 (IPv6) Addressing Architecture	7.8.1
4443	Internet Control Message Protocol (ICMPv6) for the IPv6 Specification	7.8.1
4861	Neighbor Discovery for IPv6	8.3.12.0
4862	IPv6 Stateless Address Autoconfiguration	8.3.12.0
5175	IPv6 Router Advertisement Flags Option	8.3.12.0

Border Gateway Protocol (BGP)

The following table lists the Dell Networking OS support per platform for BGP protocols.

Table 81. Border Gateway Protocol (BGP)

RFC#	Full Name	S-Series/Z-Series
1997	BGP ComAmtturnibituitees	7.8.1
2385	Protection of BGP Sessions via the TCP MD5 Signature Option	7.8.1
2439	BGP Route Flap Damping	7.8.1

RFC#	Full Name	S-Series/Z-Series
2545	Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing	
2796	BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)	7.8.1
2842	Capabilities Advertisement with BGP-4	7.8.1
2858	Multiprotocol Extensions for BGP-4	7.8.1
2918	Route Refresh Capability for BGP-4	7.8.1
3065	Autonomous System Confederations for BGP	7.8.1
4360	BGP Extended Communities Attribute	7.8.1
4893	BGP Support for Four-octet AS Number Space	7.8.1
5396	Textual Representation of Autonomous System (AS) Numbers	8.1.2
draft-ietf-idrbgp4- 20	A Border Gateway Protocol 4 (BGP-4)	7.8.1
draft-ietf-idrrestart- 06	Graceful Restart Mechanism for BGP	7.8.1

Open Shortest Path First (OSPF)

The following table lists the Dell Networking OS support per platform for OSPF protocol.

Table 82. Open Shortest Path First (OSPF)

RFC#	Full Name	S-Series/Z-Series
1587	The OSPF Not-So-Stubby Area (NSSA) Option	7.6.1
2154	OSPF with Digital Signatures	7.6.1
2328	OSPF Version 2	7.6.1
2370	The OSPF Opaque LSA Option	7.6.1
2740	OSPF for IPv6	9.1(0.0)
3623	Graceful OSPF Restart	7.8.1
4222	Prioritized Treatment of Specific OSPF Version 2 Packets and Congestion Avoidance	7.6.1

Intermediate System to Intermediate System (IS-IS)

The following table lists the Dell Networking OS support per platform for IS-IS protocol.

Table 83. Intermediate System to Intermediate System (IS-IS)

RFC#	Full Name	S-Series
1142	OSI IS-IS Intra-Domain Routing Protocol (ISO DP 10589)	
1195	Use of OSI IS-IS for Routing in TCP/IP and Dual Environments	
2763	Dynamic Hostname Exchange Mechanism for IS-IS	
2966	Domain-wide Prefix Distribution with Two-Level IS-IS	
3373	Three-Way Handshake for Intermediate System to Intermediate System (IS-IS) Point-to-Point Adjacencies	
3567	IS-IS ACruythpetongtircaapthioicn	
3784	Intermediate System to Intermediate System (IS-IS) Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)	
5120	MT-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)	
5306	Restart Signaling for IS-IS	
5308	Routing IPv6 with IS-IS	8.3.10.0
draft-ietf-isis-igpp2p- over-lan-06	Point-to-point operation over LAN in link-state routing protocols	
draft-kaplan-isis-e xt-eth-02	Extended Ethernet Frame Size Support	

Routing Information Protocol (RIP)

The following table lists the Dell Networking OS support per platform for RIP protocol.

Table 84. Routing Information Protocol (RIP)

RFC#	Full Name	S-Series
1058	Routing Information Protocol	7.8.1
2453	RIP Version	7.8.1
4191	Default Router Preferences and More-Specific Routes	8.3.12.0

Multicast

The following table lists the Dell Networking OS support per platform for Multicast protocol.

Table 85. Multicast

RFC#	Full Name	S-Series
1112	Host Extensions for IP Multicasting	7.8.1
2236	Internet Group Management Protocol, Version 2	7.8.1
2710	Multicast Listener Discovery (MLD) for IPv6	
3376	Internet Group Management Protocol, Version 3	7.8.1
3569	An Overview of Source-Specific Multicast (SSM)	7.8.1 SSM for IPv4
3618	Multicast Source Discovery Protocol (MSDP)	
3810	Multicast Listener Discovery Version 2 (MLDv2) for IPv6	
3973	Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised)	
4541	Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches	7.6.1 (IGMPv1/v2)
draft-ietf-pim -sm-v2-new- 05	Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)	7.8.1 PIM-SM for IPv4

Network Management

The following table lists the Dell Networking OS support per platform for network management protocol.

Table 86. Network Management

RFC#	Full Name	S4810	S4820T	Z-Series
1155	Structure and Identification of Management Information for TCP/IP-based Internets	7.6.1		
1156	Management Information Base for Network	7.6.1		

RFC#	Full Name	S4810	S4820T	Z-Series
	Management of TCP/IP-based internets			
1157	A Simple Network Management Protocol (SNMP)	7.6.1		
1212	Concise MIB Definitions	7.6.1		
1215	A Convention for Defining Traps for use with the SNMP	7.6.1		
1493	Definitions of Managed Objects for Bridges [except for the dot1dTpLearnedEntryDiscards object]	7.6.1		
1724	RIP Version 2 MIB Extension			
1850	OSPF Version 2 Management Information Base	7.6.1		
1901	Introduction to Community-based SNMPv2	7.6.1		
2011	SNMPv2 Management Information Base for the Internet Protocol using SMIv2	7.6.1		
2012	SNMPv2 Management Information Base for the Transmission Control Protocol using SMIv2	7.6.1		
2013	SNMPv2 Management Information Base for the User Datagram Protocol using SMIv2	7.6.1		
2024	Definitions of Managed Objects for Data Link Switching using SMIv2	7.6.1		
2096	IP Forwarding Table MIB	7.6.1		
2558	Definitions of Managed Objects for the Synchronous Optical Network/Synchronous			

RFC#	Full Name	S4810	S4820T	Z-Series
	Digital Hierarchy (SONET/ SDH) Interface Type			
2570	Introduction and Applicability Statements for Internet Standard Management Framework	7.6.1		
2571	An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks	7.6.1		
2572	Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)	7.6.1		
2574	User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)	7.6.1		
2575	View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)	7.6.1		
2576	Coexistence Between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework	7.6.1		
2578	Structure of Management Information Version 2 (SMIv2)	7.6.1		
2579	Textual Conventions for SMIv2	7.6.1		
2580	Conformance Statements for SMIv2	7.6.1		
2618	RADIUS Authentication Client MIB, except the following four counters: radiusAuthClientInvalidServerAddresses	7.6.1		

RFC#	Full Name	S4810	S4820T	Z-Series
	radiusAuthClientMalformedAccessResponses radiusAuthClientUnknownTypes radiusAuthClientPacketsDropped			
2698	A Two Rate Three Color Marker	9.5.(0.0)	9.5.(0.0)	9.5.(0.0)
3635	Definitions of Managed Objects for the Ethernet-like Interface Types	7.6.1		
2674	Definitions of Managed Objects for Bridges with Traffic Classes, Multicast Filtering and Virtual LAN Extensions	7.6.1		
2787	Definitions of Managed Objects for the Virtual Router Redundancy Protocol	7.6.1		
2819	Remote Network Monitoring Management Information Base: Ethernet Statistics Table, Ethernet History Control Table, Ethernet History Table, Alarm Table, Event Table, Log Table	7.6.1		
2863	The Interfaces Group MIB	7.6.1		
2865	Remote Authentication Dial In User Service (RADIUS)	7.6.1		
3273	Remote Network Monitoring Management Information Base for High Capacity Networks (64 bits): Ethernet Statistics High-Capacity Table, Ethernet History High-Capacity Table	7.6.1		
3416	Version 2 of the Protocol Operations for the Simple	7.6.1		

RFC#	Full Name	S4810	S4820T	Z-Series
	Network Management Protocol (SNMP)			
3418	Management Information Base (MIB) for the Simple Network Management Protocol (SNMP)	7.6.1		
3434	Remote Monitoring MIB Extensions for High Capacity Alarms, High-Capacity Alarm Table (64 bits)	7.6.1		
3580	IEEE 802.1X Remote Authentication Dial In User Service (RADIUS) Usage Guidelines	7.6.1		
3815	Definitions of Managed Objects for the Multiprotocol Label Switching (MPLS), Label Distribution Protocol (LDP)			
4001	Textual Conventions for Internet Network Addresses	8.3.12		
4292	IP Forwarding Table MIB	9.5.(0.0)	9.5.(0.0)	9.5.(0.0)
4750	OSPF Version 2 Management Information Base	9.5.(0.0)	9.5.(0.0)	9.5.(0.0)
4502	RMON v2 MIB	9.5(0.0)	9.5(0.0)	9.5(0.0)
5060	Protocol Independent Multicast MIB	7.8.1		
ANSI/TIA-1057	The LLDP Management Information Base extension module for TIA-TR41.4 Media Endpoint Discovery information	7.7.1		
draft-grant-tacacs-02	The TACACS+ Protocol	7.6.1		
draft-ietf-idr-bgp4-mib-06	Definitions of Managed Objects for the Fourth Version of the Border	7.8.1		

RFC#	Full Name	S4810	S4820T	Z-Series
	Gateway Protocol (BGP-4) using SMIv2			
draft-ietf-isis-wgmib- 16	Management Information Base for Intermediate System to Intermediate System (IS-IS): isisSysObject (top level scalar objects) isisISAdjTable isisISAdjAreaAddrTable isisISAdjIPAddrTable isisISAdjProtSuppTable			
draft-ietf-netmod-interfaces-cfg-03	Defines a YANG data model for the configuration of network interfaces. Used in the Programmatic Interface RESTAPI feature.	9.2(0.0)	9.2(0.0)	9.2(0.0)
IEEE 802.1AB	Management Information Base module for LLDP configuration, statistics, local system data and remote systems data components.	7.7.1		
IEEE 802.1AB	The LLDP Management Information Base extension module for IEEE 802.1 organizationally defined discovery information. (LLDP DOT1 MIB and LLDP DOT3 MIB)	7.7.1		
IEEE 802.1AB	The LLDP Management Information Base extension module for IEEE 802.3 organizationally defined discovery information. (LLDP DOT1 MIB and LLDP DOT3 MIB)	7.7.1		
ruzin-mstp-mib-02 (Traps)	Definitions of Managed Objects for Bridges with	7.6.1		

RFC#	Full Name	S4810	S4820T	Z-Series
	Multiple Spanning Tree Protocol			
sFlow.org	sFlow Version 5	7.7.1		
sFlow.org	sFlow Version 5 MIB	7.7.1		
FORCE10-BGP4-V2-MIB	Force10 BGP MIB (draft-ietf-idr-bgp4-mibv2-05)	7.8.1		
f10–bmp-mib	Force10 Bare Metal Provisioning MIB	9.2(0.0)	9.2.(0.0)	9.2.(0.0)
FORCE10-FIB-MIB	Force10 CIDR Multipath Routes MIB (The IP Forwarding Table provides information that you can use to determine the egress port of an IP packet and troubleshoot an IP reachability issue. It reports the autonomous system of the next hop, multiple next hop support, and policy routing support)			
FORCE10-CS-CHASSIS-MIB	Force10 C-Series Enterprise Chassis MIB			
FORCE10-IF-EXTENSION-MIB	Force10 Enterprise IF Extension MIB (extends the Interfaces portion of the MIB-2 (RFC 1213) by providing proprietary SNMP OIDs for other counters displayed in the "show interfaces" output)	7.6.1		
FORCE10-LINKAGG-MIB	Force10 Enterprise Link Aggregation MIB	7.6.1		
FORCE10-CHASSIS-MIB	Force10 E-Series Enterprise Chassis MIB			
FORCE10-COPY-CONFIG-MIB	Force10 File Copy MIB (supporting SNMP SET operation)	7.7.1		
FORCE10-MONMIB	Force10 Monitoring MIB	7.6.1		
FORCE10-PRODUCTS-MIB	Force10 Product Object Identifier MIB	7.6.1		

RFC#	Full Name	S4810	S4820T	Z-Series
FORCE10-SS-CHASSIS-MIB	Force10 S-Series Enterprise Chassis MIB	7.6.1		
FORCE10-SMI	Force10 Structure of Management Information	7.6.1		
FORCE10-SYSTEM-COMPONENT-MIB	Force10 System Component MIB (enables the user to view CAM usage information)	7.6.1		
FORCE10-TC-MIB	Force10 Textual Convention	7.6.1		
FORCE10-TRAP-ALARM-MIB	Force10 Trap Alarm MIB	7.6.1		

MIB Location

You can find Force10 MIBs under the Force10 MIBs subhead on the Documentation page of iSupport:

https://www.force10networks.com/CSPortal20/KnowledgeBase/Documentation.aspx

You also can obtain a list of selected MIBs and their OIDs at the following URL:

https://www.force10networks.com/CSPortal20/Main/Login.aspx

Some pages of iSupport require a login. To request an iSupport account, go to:

https://www.force10networks.com/CSPortal20/AccountRequest/AccountRequest.aspx

If you have forgotten or lost your account information, contact Dell TAC for assistance.